

National Chiao Tung University

EECS International Graduate Program

Thesis

蟻群傳染目標的最佳化--一種移動 Ad-hoc 網路服務探索演
算法

Ant Colony Infectious Target Optimization (ACITO) –
A service discovery algorithm for mobile ad-hoc
networks

Student: Emilio Sanchez

Advisor: Prof. Huang, Ching-Yao

August, 2014

Ant Colony Infectious Target Optimization (ACITO) – A service
discovery algorithm for mobile ad-hoc networks

研究生：桑正翰

Student: Emilio Sanchez

指導教授：黃經堯

Advisor: Huang, Ching-Yao

國立交通大學

電機資訊國際學位學程

碩士論文



Electrical Engineering and Computer Science

August 2014

Hsinchu, Taiwan, Republic of China

中華民國一百零三年八月

摘要

基於已被提出之各種不同服務探索協定的特徵，本研究專注於探討資源與服務探索協定的發展，利用網路中各個節點的數據型態，創造和主體相關的模塊。蟻群優化算法和傳染病傳播概念的優化被應用到整個網路訊息的擴散中，並在網路中開始請求搜尋的節點、與提供服務的節點之間，創造溝通橋樑。

本文中使用的�方法包括對不同產品的研究、依設計功能來做的分類、這些應用程式對應 MANET 網路的特點並且通過資源發展和服務探索協定來驗證其對事件發生的功能和適當的反應，由於網路動態的特性和選擇一個適當網路的仿真工具去複製和測試協議中的預期結果，顯示出它的優點和不足處。通過各種不同圖形的比較和相對應的描述來展現資源和服務探索協定的效率和優越表現。通過對沒有數據類型的創造或者本體論分類和訊息使用數量的相比，在協議傳播階段，消息傳播數量得到有效的降低。在網路中測試衝擊效應展示出了協議的高可靠性和在衝擊隨時可能增加爆發時，在網路中服務發現的數量產生永恆不變的結果。其他結果包括在協議中系統過載導致的無效搜索影響，搜索訊息的數量和在兼容匹配時間的影響。

在本文中發展出的資源和服務探索協定，不僅達到了 MANET 網路的特性和需求，同時通過模擬得到的結果顯示在先進工藝開發過程中 MANET 網路的多功能性、運用性、與其在更先進程序中----例如利用網路中各個不同的設備來完成的自動服務組合----的先進性。

Abstract

Nodes participating in mobile ad-hoc networks have the liberty to join, leave and move across it, inducing the constant change in the topology and presenting a challenge for the composition and maintenance of the communication via between nodes. Considering the characteristics of different service discovery protocols proposed in the past years, the present work explores the development of a resource and service discovery protocol that makes use of the data type each node in the network handles, to create blocks related among them through an ontology. Concepts of ant colony optimization algorithms and infectious spread are applied to optimize the diffusion of message across the network, as well as to create the communication links between nodes starting search requests and nodes offering services in the network.

The methods used in the thesis include the study of different works as well as the classification of the same according to its design features, the application of these solutions to fit the characteristics of MANET networks, the development of the proposed resource and service discovery protocol validating its functionalities and proper response to the events occurring due to the dynamic nature of the network and the selection of an appropriate network simulation tool to reproduce and test the expected outcomes of the proposed protocol showing the advantages and limitations of the same. The efficiency and satisfactory performance of the proposed resource and service discovery protocol is presented in the different graphical comparisons and described in the corresponding subsections of the same. A significant decrease of the number of transmitted messages during the dissemination phase of the proposed protocol is found as compared to the amount of messages used when no data type relation is created or classified through an ontology. The results obtained when testing for the miss hit effect in the network show the high level of reliability of the protocol, producing constant outcomes of the number of services discovered in the network when the ratio of invalid services populating the network increases. Other results include the invalid search effect on system overload induced by the protocol, amount of search messages and effect on compatibility match process time.

The resource and service discovery protocol developed and fully implemented throughout the present thesis not only attains to the requirements and characteristics of MANET networks, but the results obtained through the simulations show the capability, functionality and advantages of its application in the development of more advanced processes such as automatic service composition making use of data type handled by different devices in a network.

Table of Contents

Chinese Abstract	i
English Abstract	ii
Table of Contents	iii
List of Tables	v
List of Figures	vi
Symbols	vii
I.	Introduction.....	1
1.1	Motivation.....	2
1.2	Objectives.....	3
1.3	Thesis structure.....	4
II.	Mobile ad-hoc Networks – A brief introduction	6
2.1	Mobile ad-hoc Networks.....	6
2.1.1	MANET fundamentals.....	8
2.1.2	Access to communication.....	9
2.1.3	Routing protocols in MANET networks.....	10
2.1.4	MANET applications.....	11
2.2	Simulation and parameters.....	12
2.2.1	Network simulators.....	12
2.2.2	Parameters.....	14
2.3	Conclusion.....	16
III.	Current state of service discovery.....	17
3.1	Service discovery.....	17
3.2	Classification of current solutions.....	17
3.2.1	Use of directories.....	18
3.2.2	Service description.....	20
3.2.3	Response to nodes mobility.....	20
3.2.4	Selection mechanism.....	21
3.3	Conclusion.....	21
IV.	Resource and service discovery through compatibility..	23

4.1	Overview.....	23
4.2	Dissemination layer.....	25
4.2.1	Neighbor Detection.....	26
4.2.2	Reliability.....	27
4.2.3	Propagation limitation.....	28
4.2.4	Ant colony and Infectious concept for message spread...	29
4.2.5	Network events.....	31
4.2.6	Dissemination algorithm.....	32
4.2.7	Applications.....	33
4.3	Compatibility match and search mechanism.....	34
4.3.1	Service and search descriptions.....	34
4.3.2	Ontology classification.....	35
4.3.3	Search process.....	37
4.3.4	Network change adaptability.....	39
4.4	Route Composition.....	41
4.4.1	Connectivity bridge formation.....	42
4.4.2	Route control.....	44
4.4.3	Route composition algorithm.....	45
V.	Protocol Assessment.....	47
5.1	Network and environment configuration.....	47
5.2	Dissemination layer message reduction.....	50
5.3	Protocol reliability through miss hit occurrence.....	53
5.4	System overload with active protocol.....	56
VI.	Conclusion.....	58
6.1	Thesis contributions.....	58
References	59
Autobiography	63

List of Tables

Table 1	Network Simulator Characteristics.....	18
Table 2	NS-2 parameters configuration for simulation.....	20
Table 3	Service discovery protocol parameters configuration for simulation	53



List of Figures

Figure 1	Mobile ad-hoc network representation.....	13
Figure 2	Service discovery protocol process.....	29
Figure 3	Dissemination of data parameters.....	29
Figure 4	Neighbor detection process.....	30
Figure 5	Table entries for neighbor nodes in a network	31
Figure 6	Data type tables for nodes in a network	33
Figure 7	Carrier dissemination and projected number of hops in a network configuration	35
Figure 8	Data type dissemination algorithm	37
Figure 9	Representation of relations between concepts in an ontology	40
Figure 10	Search mechanism adaptability to network change and cancellation request process	44
Figure 11	Link table and bridge formation in mobile ad hoc configuration	46
Figure 12	Bridge elimination after lost connectivity between two nodes in the network	48
Figure 13	Route composition algorithm	50
Figure 14	Simulation environment for service discovery protocol	52
Figure 15	Reduction of the amount of transferred messages in the dissemination process	55
Figure 16	Average time for request/service compatibility match in search process	56
Figure 17	Invalid search response effect in service discovery protocol	58
Figure 18	Invalid search response effect in compatibility match process time	59
Figure 19	Invalid search response effect in number of search messages disseminated	60
Figure 20	Invalid search response effect in system overload with active protocol	61

Symbols

TTL: Time-to-live

t_B : Timed beacons

T_w : Waiting period

T_t : Transmission time in medium

ACK: Acknowledgement

O : Ontology

H_p : Projected total number of hops

H_{max} : Maximum number of hops

nID : Node identifier

H_{eff} : Projected hops effective value

R_S : Request search message

R_C : Request cancellation message

R_R : Request response message

ID_R : Search request identifier

D_S : Service description

LT: Link Table

B : Bridge (LT entry)

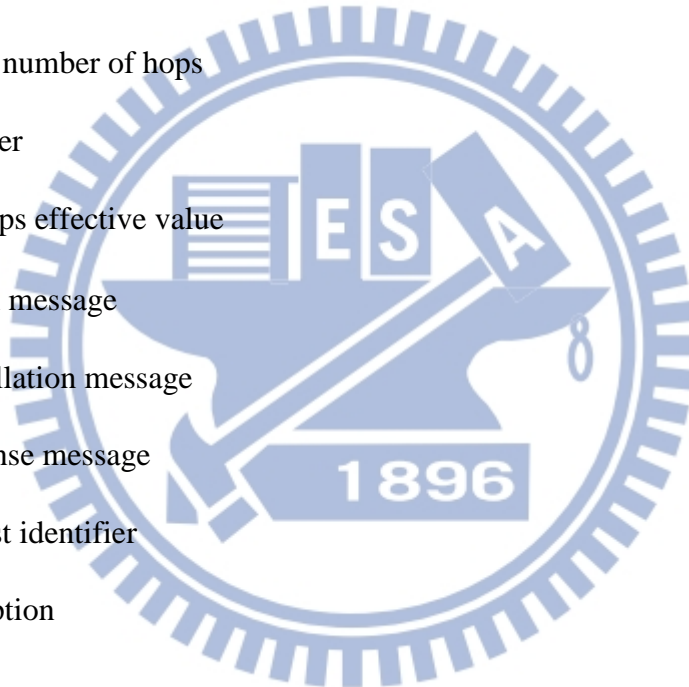
ID_B : Bridge identifier

m_{elim} : Elimination message

A_{spHops} : Average shortest-path hop count

ANP: Average amount of network partitioning

FRR: False response ratio



I. Introduction

In the ubiquitous computing world, users should not be limited to the services and resources offered by their individual devices, but be able to access the services and resources of all available interconnected machines to realize complex tasks and functions, enhancing the quality of results obtained. At the beginning when computers were first interconnected between each other, made possible to share computational resources and information inside universities and research facilities. To achieve this capability, years of study and effort were spent to resolve the many different problems associated in this field, communication capability, reliability and error control are just some worth to mention.

The wireless communication, downsizing of electronics and increasing demand of mobile device ownership by the “always connected” generation of people had pushed technology to become more accessible, providing a very high mobility and freedom for users to access these types of interconnected networks moving towards the birth and development of the “Internet of Things”.

According to the ubiquitous computing model presented by Mark Weiser [1] , the distributed computing was not limited to a single computer or a determined network of machines, but could be formed by different relatively complex devices spread in the user’s surrounding but integrated in the network. To reach the objectives of ubiquitous computing, the devices need to be aware of other devices in the network and capable to collaborate among them to try and solve a user’s problem. The user simply makes use of the computer potential without being concerned of the other devices in the surrounding or the interconnection needed between them to perform a task, hence the user should have a minimum interaction role with the process of connection.

Complex services are formed by the interconnection of simpler services, joining their individual capabilities to perform more sophisticated processes, this is called the Service Composition process and many researchers chose this field for the potential and application in the Internet of Things, as well as the challenges to be addressed in the area. [2] introduces a task computing process, where the user specifies a task to be solved and the system executes the task using the resources and services available. One of the most important challenge that require attention for the Service Composition process is the ability to create awareness of the resources and services available in the environment, without a proper knowledge of the resources a device can access, the service composition will not be capable of providing the efficiency, dynamicity

and quality of service to the user, hence; throughout the years, Resource & Service Discovery algorithms had been developed and researched to improve the service discovery process while attaining to the limitations and mobility of devices in the network.

The development of the semantic web [3], opened new trends for the global computing vision, proposing the use of semantic descriptions to record resources, extract the meaning of the available information automatically and build decisions upon it. Ontologies are used for these purposes, providing a vocabulary of classes and relations that allow to describe a domain. Through ontologies, a system is able to recognize relations between the different terms and make use of them to process information. The semantic description is applicable to any resources such as videos, images, text and also to services provided by the devices in a network.

The different topologies of the network carry their own requirements and limitations for the system, hence is important to realize the different characteristics concerning mobility, communication mechanism and available computational resources. The current work is developed for networks constituted spontaneously by the devices that form it, commonly known as Mobile Ad hoc Networks, this type of networks do not have a fixed infrastructure for the communication among agents, the different devices that constitute the network participate actively in the communication through message routing and message hopping to reach other agents and start the service discovery process in a dynamic manner.

In the early nineties, the first ant colony optimization algorithm was proposed, according to [4], *“ants have inspired a number of methods and techniques among which the most studied and the most successful is the general purpose optimization technique known as ant colony optimization.”* Using some characteristics of the ant colony optimization algorithm and concepts of an infectious behavior, the developed solution aims to create an efficient service discovery algorithm that takes advantage of service compatibility among agents using the ontology of concepts to organize the comparable parameters.

The objective of the thesis is to develop and present the resource and service discovery algorithm named Ant Colony Infectious Target Optimization or ACITO, the functionalities, efficiency, advantages and limitations for service discovery in mobile ad hoc networks.

1.1 Motivation

Resource and service discovery protocols had been topics of interest for many researchers and academics, the different considerations and characteristics of networks where a particular

protocol will be implemented vary in a wide range of requirements, limitations and challenges. Mobile ad-hoc networks or MANET networks are no different, and the absence of a fixed topology together with the relative high mobility of nodes participating in the network has shown to be a problematic that requires a well-designed and robust solution to properly operate in the conditions of this type of networks.

The modern world is becoming increasingly dependent on technology, gadgets, devices and other technologies are present and accompanying the majority of people's daily routines. Repetitive tasks are solved automatically, while more complex routines are starting to require just a minimum intervention from the user to produce the expected results.

The constant search for automatized systems relies on the interconnection between these devices to overcome the barriers imposed by the limitations of an individual device but rather take advantage of all the resources and services available in the surroundings of the machine. This is the main objective of a resource and service discovery protocol, to efficiently locate the resources and services available in the network.

Current solutions make use of different factors to create the most suitable method to fit the characteristics of a particular network, works stated in the reference require either a global knowledge of all the nodes participating in the network at a moment, or the proposed solutions still rely heavily on user participation. The objective of the present work is to present a service discovery protocol that uses a semantic classification of the different types of data handle by the nodes in a network as a way to categorize, match requests and services and allow an efficient dissemination of available services in the network attaining to the requirements and limitations of the mobile ad-hoc networks.

Semantic technologies had been used in previous works [5] [6] , the classification of services in these works is performed in a general manner or using identifiers as a mean of classify the services.

1.2 Objectives

To achieve the development of the protocol with its functionalities and characteristics, a series of objectives is established:

- Considering the distributed nature of the MANET networks, define the dissemination process that will allow nodes to announce and promote their services to other nodes

participating in the network, using a proper service description of data type classify through an ontology.

- Development of a compatibility match oriented search mechanism that uses the data type of search requests and services to compare, analyze and match requests' descriptions with available services in the network.
- Considering the mobility of nodes in the MANET networks, define a route composition mechanism capable of creating and maintaining the connectivity between nodes participating in the network and capable of adapting to the topology changes produced by the dynamic nature of the network.
- Complete integration of the different parts that make up the full service discovery protocol and the assessment of the same through a proper network simulator to evaluate the advantages and limitations of the proposed solution.

1.3 Thesis structure

Chapter 2 presents a brief introduction of the mobile ad-hoc networks, the characteristics of the same become important when designing a protocol to fit its requirements. The different sections in the chapter present the fundamentals, applications and different types of routing protocols used in the same, as well as analyzing the network simulators and the parameters used to test the solutions.

Chapter 3 introduces a more in-depth view of the current state of service discovery containing the different classifications of solutions and subsections describing the characteristics of each type before concluding the chapter.

Chapter 4 describes the proposed resource and service discovery protocol in full and the compatibility match used in the dissemination and search processes of the same. It consists of different sections each concentrating in the different parts of the solution as well as subsections containing the details, characteristics and concepts taken into account in the design of the different parts that make the full proposed solution.

Chapter 5 shows the assessment of the proposed solution, the different sections present the configuration of the environment and the network, the different features tested during the simulation and the graphical results obtained.

To finalize, chapter 6 presents the conclusion taken after realizing the research work as well as the contributions of the thesis. Throughout the work, different chapters provide comments and

recommendations on different improvements that can be considered in future developments of the protocol.



II. Mobile ad-hoc Networks – A brief introduction

Networks can have different structures depending on the technology used for the communication of the devices, protocols and communication services. The present work focuses exclusively on the mobile ad hoc networks or wireless ad hoc networks which main feature is the absence of cables in the communication and changes in the topology of the network as a response to the mobility of the devices that form it. Section 2.1 introduces the fundamental characteristics, limitations and requirements of this type of networks, offering a general overview of the concepts applied and used throughout the remaining parts of the document. Section 2.2 introduces the parameters used in simulations as well as providing an insight of the advantages of using simulators to test and experiment with mobile ad hoc networks.

2.1 Mobile ad-hoc Networks

Mobile ad hoc networks or also known as MANET, are a type of wireless networks without a fixed infrastructure capable of re-configuring itself continuously as a response of the free independent movement of the devices present in the network. The communication in this type of networks occurs in a non-predefined manner, completely depending on the distribution of devices at a certain time, there are no dedicated agents performing communication tasks at any moment, hence message dissemination becomes a distributed task among all participant agents in the network. More detailed information about ad hoc wireless networks can be found in [7] . Protocols designed for MANET networks have to consider a set of characteristics particular to this kind of networks:

- Distributed Operation

As opposed from fixed infrastructure networks, the agents in MANET networks cannot rely on dedicated agents to provide routing for communication functionality, there are no dedicated agents in the network, hence the functionality is a result of tasks distributed among all agents in the network.

- Dynamic topology

Due to the mobility capability of agents in the network, random connection and disconnection or aleatory position change of agents within the network are common events. The protocols should perform the necessary tasks to guarantee the connectivity between agents and respond to unexpected actions.

- Fluctuations in connection capability

The communication is performed through hops of messages among the agents and in a wireless manner, hence, errors due to connectivity fluctuation are frequent, the protocols designed for this networks should control and in best effort minimize the amount of errors.

- Low power requirement

The devices populating the MANET networks are usually operating with batteries, which imposes a series of limitations to CPU, memory, communication and other resources of the devices. The protocols should take into account the limitations and consume the minimum amount of resources possible, poor design of protocols will result in complete drain of valuable resources.

When devices are positioned within a range of communication from each other, the MANET networks are formed spontaneously, through a direct connection between two or more devices in the network, the same can transmit information and share resources to solve complex tasks. On the other hand, limiting the communication to devices directly connected to each other would limit the capabilities of this type of networks, hence the ad hoc networks allow for the indirect communication between devices that are not positioned in the nearby surrounding of communication area.

To achieve this capability, the information is segregated using multiple consecutive hops from device to device until reaching the final destination. The process to create and maintain the necessary routes for this indirect communication requires the participation of different devices or agents populating the network, forming a sophisticated process that needs to abide to the characteristics and limitations mentioned before.

The agents that are part of the network are free to move in any direction while participating in the network communication, making mobility a fundamental characteristic of the MANET networks, the communication elements such as routers, bridges, etc. is explicitly separated from the computation devices such as PCs, servers, etc. Hence the mobility in fixed infrastructure networks is not a main concern since is applied to the computation devices only and does not affect the communication of the network entirely, for example in a scenario where a PC moves in the network, only the communication of other devices that want to access the resources of

the PC will be affected, but the capability of other devices to communicate with each other is not affected. In MANET networks, the differentiation between communication elements and computation devices is very ambiguous, hence the mobility of an agent in the network can and does affect critically the communication among devices.

2.1.1 MANET fundamentals

Different protocols are at work for any communication network to be able to offer functionalities to users, and the MANET networks have a strict set of requirements and characteristics for protocol designs, each with the aim of achieving a specific step that allows the next ones to function efficiently to reach the final result.

To formulate the different terms commonly used in ad hoc networks and will be adopted throughout the research, definitions and concepts are introduced as follow:

- Node
Each of the devices participating in the network, agents with computational capabilities with diverse characteristics but a common wireless communication technology that allows it to be part of the mobile ad hoc network.
- Link
Assuming the existence of wireless communication technology in two or more nodes positioned at a distance from each other, a link is defined as the interconnection or communication between the nodes.
- Neighbor
Nodes located at a close range of communication from each other, a general term would be nodes located at a distance of jumps (n) from each other capable of direct communication among neighbors of degree n .
- Distance
Refers to the amount of links that need to be created for the information to be able to reach the destination, measured in number of *jumps*.

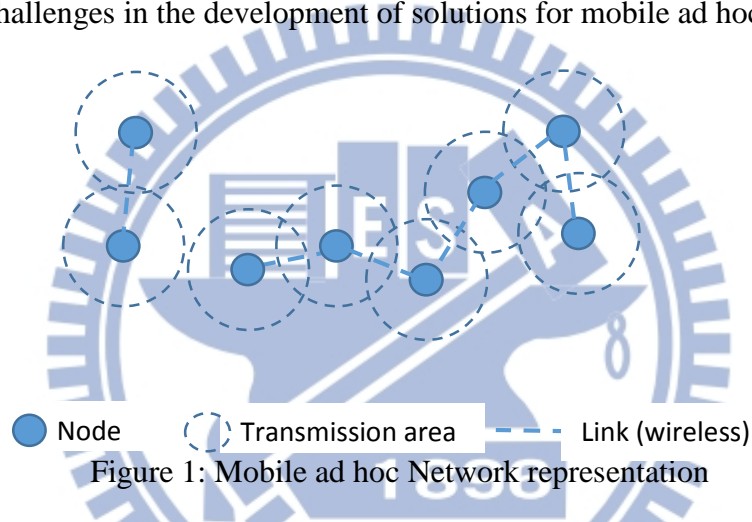
- Message

The information transmitted among nodes through the different links, the size of message and transmission speed depends on the characteristics of the wireless technology used by nodes.

- Routing

Refers to the paths through which nodes can communicate using a single or series of messages sent through different links or intermediate nodes. Due to the mobility of nodes in the network, the paths can change, appear or disappear.

The figure below shows the elements of an ad hoc network through different nodes and the communication links in between. Security, quality of service, power consumption management are all current challenges in the development of solutions for mobile ad hoc networks.



2.1.2 Access to communication

Wireless nature is an essential part of ad hoc networks and there are no dedicated transmission channels for the communication between nodes. Electromagnetic waves are propagated in the surroundings with the objective of creating the communication among nodes in the network, the absence of dedicated channels leave the waves open to elements capable of causing problems such as attenuation, interferences or even blocking of the signals.

These problems impose limitations that require especial control mechanisms to help reduce the impact of the same in the transmission process. The attenuation is the relation between the transmitted and received power, in a free space transmission scenario follows the inverse-square law which states that the received power of transmitted electromagnetic waves is proportional to the inverse of the square of the distance from the transmitted source.

$$Intensity \propto \frac{1}{distance^2} \tag{1}$$

The presence of different obstacles and materials in the environment deteriorates the received power even more, requiring a higher transmission to overcome the loss.

At the same time, there is a very high number of different transmissions active in the environment sharing the same transmission medium, the different devices with wireless transmission capability are governed by different techniques to control the access to the shared channel and this way avoid the loss of information if they all were to access the channel at the same time.

In real world application, the devices that make part of the network are handle around every day common objects (e.g., furniture, walls, windows, etc.) that present obstacles for the arriving signal, not only signal block can occur, but other phenomena such as reflection, diffraction, dispersion are all a reality cause by the intermediate physical objects.

A number of techniques and methods had been developed to reduce the negative impacts of the different problems and make a best effort to solve them. Medium Access Control (i.e., MAC) is one of the techniques developed that aims to control the time when a device can transmit using the shared channel, and is defined by the current protocol in use Wi-Fi IEEE 802.11 [8] for device communication in ad hoc networks, same as the one used for the development of the current thesis.

2.1.3 Routing protocols in MANET networks

Two groups can be defined as follow:

- Pro-actives

Different works were presented for this type of protocols and their functionality showed to be adequate for networks with a reduced number of nodes. The nodes in the network maintain a table with the information of the existing routes beforehand, a certain degree of previous knowledge is present since the information is created even before is needed, this characteristic makes it not suitable for large networks because the maintenance of the information and the constant sending of update messages create a high overhead in the system. Examples of this kind of protocols were presented in Destination Sequenced Distance-Vector [9] and more recently in Optimized Link State Routing [10] .

- Reactives

As the name implies, this type of protocols search for the communication routes at the time a node wants to establish communication with another node. Using the flooding approach, the search floods the network with the search query or message to establish a communication route, the route can be kept alive through different mechanisms for a certain period of time to overcome the changes in the topology of the ad hoc network as presented in the work [11] .

Besides the two above mentioned groups, there are a variety of other protocols that make use of different characteristics and take advantage of the different positive features of the approaches, the hybrid solutions present a combination of pro-active and reactive characteristics, others focus on location-based approaches, using the geographical position through GPS to send the messages to the closest location to the destination that needs to be reached, energy aware approaches distribute the routing load among all nodes in an evenly manner and the multicast protocols that create and arrange groups to perform the routing tasks.

2.1.4 MANET applications

The characteristics of ad hoc networks make it suitable for situations where a rapid deployment of a network is needed and previous communication infrastructure does not exist. Emergency response systems, military operations, sensor data collection and monitoring represent areas of application for ad hoc networks. With the increasing popularity of sensors usage, governments, companies and privates are increasingly interest in benefiting from the advantages of data collection, measurements and monitoring of goods and spaces. Smart cities look to provide efficient services while saving valuable resources such as energy, reduction of CO2 emissions, waste management and others, the agricultural sector relies on technology to monitor, improve and maximize crop production, forests are equipped with a number of sensors to prevent wildfires or control the wildlife well-being as well as study behaviors and changes of the same.

In the case of emergency response, many natural disasters of high scale (e.g., earthquakes, tsunamis, flooding, etc.) leave the communication infrastructures completely useless, ad hoc networks capabilities allows rescue teams to deploy communication networks formed by the devices with computational capabilities itself, aiding in the search, repair and recovery of resources made unavailable following the collapse of the traditional communication network.

2.2 Simulation and parameters

Due to the dynamic nature of MANET networks, is difficult to design and develop tests in real world scenario, a number of devices with wireless capabilities populating the network should move along an environment with variable dimensions following patterns and different range of velocity. Hence is very common the use of simulators in researches related to mobile ad hoc networks due to the capability of providing behavior models for the elements that constitute the network, characteristics such as transmission, routing protocols, message traffic and mobility are all handle in a controlled test environment.

Throughout the years, scientists and academics have been discussing the reliability and credibility of the use of network simulators, [12] and [13] consider the tests made using network simulators provide significant validation of a particular proposal, as long as the parameters used during the tests are correct and adequate to the requirements of the system. Although the most reliable results will only be obtained through real world experiments, the high inversion of capital to design the scenario and acquire the necessary devices for testing, besides the amount of time required to reach the optimum settings and without the guarantee the expected result will be reached, make the real world experimentation impractical and high risk.

The simulations performed in the research and presented in chapter 5 and 6 of the document, takes advantage of the capabilities offered by network simulators to provide a controllable environment. A range of 20 to 200 services is used in an area sized 800m X 800m, the nodes move at different speed and the response of the protocol is tested through different experiments with the adequate parameters settings. A real world experiment would require a large investment in capital and effort, factors that make it not possible to be developed for this research.

2.2.1 Network simulators

A variety of network simulators had been developed and used in many previous works related to the field. GloMoSim [14] , OPNET [15] [16] and ns-2 [17] share the common characteristic of being based on discrete event simulations. Meaning the behavior of the nodes in the network is modeled through the process of events occurring in the network, in the order they occur. The events are a response of the successes in the ad hoc network, such as sending message, receiving message, message collision; these events can also be generated by the user

through the scripts that define the behavior of the simulation. Besides, the simulators also provide models for the physical layer transmission, access and routing protocols, network traffic and mechanisms to generate and simulate the mobility of nodes inside the network.

- GloMoSim

Is a network simulator developed by the UCLA Parallel Computing Laboratory, it was developed as a wireless network simulator exclusively, the main advantage is its capability of executing in a parallel manner in different processors, reducing the execution time of a simulation. The simulator uses a parallel programming language called Parsec, and all the extensions and protocols to be included in the test environment should be written using this language. The company working on GloMoSim released the last version in the year 2000 before stop working on the open version of the software, later in 2011, a new commercially available simulator QualNet [18] was released based on the previous GloMoSim simulator characteristics.

- OPNET

A commercially available network simulator developed by OPNET Technologies with the main advantage of integrating the modeling, simulation and network analysis processes, the process is performed in a graphical approach, making it very intuitive with a low learning curve, very popular in corporation use and educational application.

- ns-2

This is probably the most well documented network simulator since is an open source software and developed in a collaborative manner, the license is distributed under GNU GPLv2 and is widely used for educational and research purposes. Currently includes contributions from different research centers and universities, making it an ideal candidate for the experimentation of protocols related to ad hoc networks. The protocols and extensions are written in C++, while the configuration of the environment and events is done using Tcl language. The main drawback with ns-2 simulator is the scalability, when a large network is simulated, the necessary computational resources increases greatly. A new version of the simulator, NS-3 [19] is being developed to facilitate the use and maintenance of the system as well as the future extensions.

Table 1: Network simulator characteristics

	GloMoSim	OPNET	ns-2
<i>Language</i>	C / Parsec	C++	C++
<i>Environment</i>	Parsec	Graphical Interface	Tcl
<i>Java Client</i>	None	None	Agent J
<i>License</i>	Educational purpose	Limited edition	GNU GPLv2
<i>Advantages</i>	Parallel execution	Simple to use	Widely use and documented
<i>Drawbacks</i>	No Java client, Parsec language	No Java client	Scalability

The ns-2 network simulator was chosen to develop the simulations required by the thesis and validate the proposal of the same through the results and comparison. The open source nature of this simulator allows for a faster and easier understanding of the functionalities and protocol extensions through the extensive documentation available. Through the Agent J extension for the java client execution, the standard libraries and UDP sockets as the communication mechanism, validates the resource and service discovery algorithm presented in chapter 5 through real implementation without the necessity of real device deployment, the algorithm was developed and tested using Java programming language and C++.

2.2.2 Parameters

To configure the test environment of a network simulator a number of parameters need to be set. The variation of the same have a drastic impact in the results obtained during the execution of the simulation.

Parameters related to the simulation tool, the environment characteristics and solution related can be distinguished while designing the solution. A great number of researches do not specify the configuration of the parameters followed during the experimentation, which makes it difficult to re-create the behavior and results for comparison between solutions.

[20] identifies different settings that need to be correctly configured in the simulator to obtain reliable results, type of antenna, transmission distance, MAC protocol, query queue management and propagation model. NS-2 specification indicates more accurate results when 2-Ray Ground propagation model, consisting on the use of a direct path and a reflected from the ground, is chosen for long range transmissions, as well as an Omni-directional antenna type.

Other parameters related to the simulation environment include the amount of nodes, mobility and topology of the environment. The mobility of nodes is limited to the topology and size of the scenario, where transmission T in an rectangular test area of dimensions (a,b) define the maximum amount of nodes n_{max} through the equation:

$$n_{max} = \frac{\sqrt{a^2 + b^2}}{T} \quad (2)$$

[21] presents a proposal for constructing scenarios that allow for a more rigorous evaluation of MANET networks solutions. To perform the simulations in this research, the two metrics proposed by the work were selected and are described below:

- Average network partition

Refers to the average percentage of nodes in a network that are disconnected in a given period. If the value is high, the functionality of a protocol for mobile ad hoc networks will not be reliable, low values for this parameter guarantees the protocol is creating communication routes among the majority of nodes in a network.

- Average shortest path

Refers to the average shortest distance allowed between two nodes in the network. Incorrect settings of parameters may create a scenario where a great number of nodes are located at a one hop distance from each other, creating wrong measurements. The average number of jumps necessary for messages to reach a relevant node in the network should be well configured.

[22] identifies some mobility models that are significant to the evaluation of protocols. The selection of a mobility model depends on the concentration of the application for the developed protocol. For the experimentation in the thesis, the Random Waypoint model was chosen as the mobility model, based on the movement of nodes with aleatory speeds and directions it provides a general overview of the characteristics of the proposed solution without the need of moving into more specific details of the application.

Standards for the evaluation of the resource and service discovery protocols such as the one presented in this work are difficult to find, the test configurations and different proposals comparison are limited, a best effort assessment and analogy was applied during the test phase.

The table below shows the different parameters common for the experimentation presented in chapter 5, and that were selected and set following the typical configurations for mobile ad hoc networks.

Table 2: NS-2 parameters configuration for simulation

Simulation Parameter	Value
Type of antenna	Omni-directional
Propagation model	2-Ray Ground
MAC protocol	IEEE 802.11
Query queue	5 max.
Transmission range	100m
Transmission speed	[11 , 54] Mb/s
UDP packet (max.)	1500 bytes

2.3 Conclusion

The chapter introduced the fundamental characteristics of MANET networks to provide a better understanding of the limitations and requirements of this type of networks as well as the applicable solutions that fit the wireless transmission, access control and routing protocols of the same.

A comparison of the different network simulators available is presented, and the characteristics and advantages of the chosen ns-2 simulator along with a reasoning to be the simulation tool for the present work is also introduced. The possibility of performing simulations using the Java client and communication libraries fit the requirements to evaluate the proposed protocol.

The different parameters that act and affect the simulation and validation of the same where introduced and discussed in the final part, along with the set of parameters selected for the experimentation of the proposed solution in this thesis.

III. Current state of service discovery

The chapter has the objective of introducing the different solutions and methods for service discovery currently functioning in the MANET networks field. The study of the previous works and researches provides a classification of the advantages, limitations and characteristics of the solutions. The investigations showcasing the features that want to be highlighted for the present work and the ones best represented in the area were selected in the reference examination process.

3.1 Service discovery

Service discovery is the process that allows the participants of a network to realize the different services available in their surrounding as well as obtaining the necessary information for posterior access and use of the services. There are two essential processes that can be distinguished in service discovery:

- Service publish
The participants offer their services to other nodes using a publish method announcing the characteristics of the services available.
- Search request
Potential clients looking for services send a search request with the characteristics of the service that will satisfy the requirements of the search.

As mentioned in the previous chapter, there is a blurred line between service providers and clients in the MANET networks, since the participants in this type of networks usually take both roles simultaneously.

3.2 Classification of current solutions

Service discovery problem in ad hoc networks gathered great attention from researchers, many solutions try to adequate the traditional infrastructure solutions to the characteristics of mobile networks. [23] proposes a series of aspects considered to present a classification of the solutions.

The following sections of the chapter present details of the aspects and classification of the solutions relevant to the present work:

3.2.1 Use of directories

The category includes two different proposals, solutions that make use of dedicated nodes that hold the directories containing the information of the services available in the network, and the ones that are based on the dissemination of the information among nodes without the use of dedicated nodes as directories holder.

The use of directories had been used since the first solutions for service discovery came out, where service providers would register their services through an administrator in one or many dedicated machines in the network, and the clients would search in the dedicated machines following the same process. For networks where there is a designated infrastructure, the use of directories represents a very efficient approach, since the mobility is limited to a few number of participants in the network and the same does not affect the communication between other nodes.

The characteristics of MANET networks does not allow for this approach to be viable for various reasons, the first one being the necessity of the services directories be located in nodes that are part of the ad hoc networks since there are no dedicated infrastructure elements specifically designed for this purpose. Another factor is the mobility of nodes in this type of networks, the communication routes will change due to the mobility of nodes forming the network, hence the service directories contained in a node may suddenly become unavailable; the communication routes also present another problem, since the routes from a node to another are rarely direct, the messages require multiple jumps between different nodes of the network.

To provide a suitable solution for the use of directories in ad hoc networks, a mechanism capable of dynamic selection of suitable nodes to take the role of service directories able to react to network changes is proposed in the work of [24] . Due to the distributed nature of ad hoc networks, the directories communicate with each other, allowing for information exchange and open access to services available in different areas of the network.

The work in [25] proposes the construction of a minimum dominant set among all the nodes in a network, where the selection of directories has the objective that all member of the network are covered for at least one of the directories. The information interchange among the directories is done using Bloom filters [26] to re-route the search requests to the directories that contain the services being searched.

Creating groups of nodes that have relation with each other is another approach taken for the use of directories in ad hoc networks. [27] proposed creating rings of service providers, these rings were created with providers located physically nearby or providing related services. Each group contains an access point to a node selected using a specialized algorithm, this access point provides the means to send search requests to different set of rings across the network.

In 2011, [28] combines the grouping concept and use of directories with GPS location of the nodes participating in a network to create regions with the geographic characteristics of the nodes. The work in [29] make use of distributed hash tables – DHT in selected nodes of the network to distribute the information among all the participating nodes. The directories are updated with the information coming from other nodes and at the same time the requests are redirected using the distributed keys, eliminating the unnecessary duplication of the information in all directories of the network.

On the other hand, the network dissemination are based on the diffusion of service providers and client requests throughout all the nodes of the network. In this way eliminating the need of centralized service information in a set of selected directories. The challenges in this type of approach include the frequency of diffusion, propagation distance of the messages and the method to be used for the same. [30] introduces the service information dissemination through multiple jumps among nodes in the network, each node propagates the information to its neighbors located at a one jump distance from each other and the frequency of propagation follows an exponential decay. Later works made use of the concepts proposed in [30] and combines it with the capabilities of the devices participating in a network (e.g., bandwidth, power, etc.).

[5] makes use of counters called Time To Live – TTL, to control the distance for the dissemination of service information with counters decreasing with each jump, the nodes form groups capable of saving the information provided by other nodes in the network, eventually any node is capable of responding to any incoming request. Another approach consists on the selective re-sending of messages, where the nodes that receive a request determine to which concrete neighbor the message should be re-sent, [31] makes use of this concept using the geographical information of the nodes through GPS. A more recent work [32] explains a mechanism where intermediate nodes receive the information published by service providers, and store it to later be able to answer to a request from a client if the search is compatible.

3.2.2 Service description

The techniques apply to determine the compatibility between the published service information and the search request can be syntactic or semantic. Syntactic solutions make use of the key-value pair to establish compatibility while the semantic solutions make use of advanced models such as the use of ontologies.

A solution has greater expressivity when the technique applied for compatibility pairing of services and requests allows for easy description understandable by both parties and with no ambiguity that could cause an error. Various works make use of Unique Universal Identifiers – UUID to localize elements that have an exact match in the requests, UUIDs size are limited to few bytes, making them suitable for communication mechanisms with limited capabilities as presented by [23] . The main drawback is the necessary previous knowledge required for these identifiers, making it a not flexible solution for ad hoc networks.

Key-value pairs allow for a greater expression during the search process, the attributes of the service are described in any text format, although the most common formats are XML or WSDL for their characteristics as service description languages. [33] argue the limitations of the syntactic approach, the system is limited to a set of keywords without being able to perform a difference/similarity comparison and often returning an empty answer for the request. Ontologies become an alternative to overcome the syntactic approach limitations, although the nodes participating in the network should have a previous knowledge of the ontology being used a greater expression in search descriptions is achieved, various works make use of ontologies for the classification, comparison and compatibility match among service search and description. [6] [34] .

3.2.3 Response to nodes mobility

The nodes participating in an ad hoc network are free to change positions in any direction while populating the network, this implies the services being discovered in the network need a mechanism to maintain the updates and connectivity to avoid the inaccessibility of services in the dynamic network.

The work in [35] make use of an active probe mechanism, where requests are periodically sent to previously discovered services and this way determine if the service is still available or has been lost. On the other hand, the service providers can make use of notifications containing information of unavailability or modification of a service, to alert nodes interest in the

mentioned service. This approach has the advantage of providing a more immediate response to changes in a particular service as presented in [36] .

If the amount of nodes in a network is very high, the increasing service search requests active in the network can create a high overhead in the system, affecting directly the active probe mechanisms and notifications mentioned previously. [37] also mentions the considerations for approaches relying on the creation of groups of nodes, as mentioned in section 3.2.1, a mechanism to refresh the information necessary for the creation of the groups needs to respond to changes in the network as well as necessary updates.

3.2.4 Selection mechanism

In the service discovery process, most probably many responses will be produced for a single search request depending on the specific or generic nature of the description used in the request. Hence, the selection mechanism has the objective of selecting the most suitable service among all the options received for a relevant search.

The work in [23] classifies the selections into two separate categories:

- Automatic
Specialized algorithms use different criteria to decide the most suitable service for a specific search request. [38] considers the number of jumps and service capabilities as the criteria to determine the most suitable response for a search.
- Assisted
Where the user is presented with a set of services that fulfill the search requirements and is able to choose independently.

3.3 Conclusion

A study of the current state of service discovery was introduced and analyzed in depth to gain a better understanding of the characteristics, the features, classification and concepts applied by researchers and the methods developed with their advantages and limitations that can aid in the improvement of the proposal in the present paper.

Due to the various factors that need attention in MANET networks, as well as the requirements and characteristics to be considered when designing and evaluating protocols directed to this kind of networks, there is no clear agreement about which approach is better than other. The target environment for a specific protocol can vary resulting in different choices to reach the

desired outcome, as mentioned in section 3.2.1, the use of directories would not fit a network populated by nodes with high mobility due to the constant need of updates and unnecessary duplication of information.

The proposed service discovery protocol in chapter 5 makes use of the dissemination approach, no directories with centralized information of the services is used and multiple jumps between nodes is selected for reasons that will be discussed in the chapter with the advantages and relevant functionalities applicable to MANET networks.



IV. Resource and service discovery through compatibility

The chapter introduces the proposed discovery protocol built based on the compatibility of service parameters. Using ant colony and infectious spread concepts, the dissemination layer aims to transmit the different types of input and output parameters classified through an ontology of concepts. The relation between the parameters allows to create a set of compatible data between nodes that can be used to optimize the service search process by decreasing the amount of messages necessary to be sent through the network while allowing considerable high level of expression for service descriptions in MANET networks.

The contributions of the present work include a dissemination layer using ant colony and infectious spread characteristics to extend parameters of services across the different participant nodes in the network, creating a compatibility table organized by an ontology of concepts and route composition to provide the connectivity among nodes.

The search mechanism makes use of the compatibility of parameters from the different services to group the related nodes and improve the efficiency, referring directly to the characteristics of the input and output parameters, it finds the most suitable service for a certain search request.

To establish the connectivity between different nodes in the network when a response for the search request is received, efficiently maintaining the communication requires the creation of dynamic routes to handle the problems cause by the mobility of nodes. The routing mechanism is integrated in the protocol to aid in the dissemination and search process. As discussed in section 2.2, the proposed protocol is assessed through the use of simulators, and the comparison, advantages and limitations of the same are presented through the evaluation in chapter 5.

4.1 Overview

Service discovery algorithms main objective is to efficiently discover the services that are available in the vicinity of a particular device and build the communication route that allows the connection between devices to access the different services or be part of the middle step to achieve the composition of multiple services.

As discussed in section 3.3, although there are various proposals in the current state of the solutions, there is no formal agreement to decide which one is better than other; this is due to the different requirements that need to fulfill independently by the solutions and the result that seems most optimal for the final objectives of the solution.

The main characteristics of the proposed protocol in the thesis are the compatibility considerations among the different types of data the agents or nodes handle in an environment, the single road discovery and route composition using ant colony based concepts with a set of functionalities to decrease the amount of message distribution during the dissemination phase and rapid route composition capabilities and the use of ontologies to classify the data parameters to build the compatibility match in the protocol allows for an improved response of the system to the different search queries generated in the environment.

The different state of the art solutions discussed in section 3.2 based the methods using a range of different service identifiers, creating categories of services, types, and names or using the common service identifiers.

The proposed data compatibility approach takes advantage of the type of data the nodes handle to improve the discovery process, advantages of using the data compatibility method can be seen below:

- The type of data handled by the different nodes has a direct relation to the tasks able to be performed, as well as the capabilities to fully fit the requirements of a search request, allowing also for better expression in the search description.
- As opposed to the use of identifiers, where the nodes in an environment should be aware of the particular identifier of all the nodes that participate in the network, using the data parameters classified in an ontology only requires the knowledge of the ontology itself. Allowing to establish multiple service compatibilities by reusing the descriptive types.
- The controlled nature of the protocol allows for a regulated spread during the dissemination process, avoiding the creation of system overhead and exhausting valuable node resources, making it a suitable solution for MANET networks providing a competent response of the system.

The proposed protocol is formed by three main parts each in charge of a specific task to fully fit the requirements of service discovery algorithms for MANET networks, the dissemination layer, the compatibility match search mechanism and the route composition.

The dissemination layer has the objective of letting other nodes be aware of the type of data a certain node is able to handle, it achieves this task by announcing the services that are available to be accessed as the first step to move into the compatibility match search mechanism that finds the service that fulfills correctly a search description initiated by a node in the network.

Route composition is the second action taking place, with the aim to create routes that will allow two or more relevant nodes, previously matched, to communicate with each other and access the services that are of interest. In each of the phases, there are different considerations and factors to achieve each independent objective, figure 2 shows the process and features inside the different stages of the proposed solution.

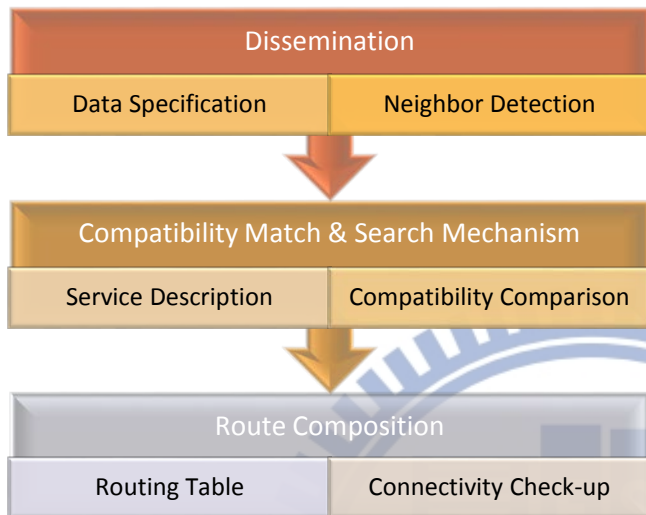


Figure 2: Service discovery protocol process

4.2 Dissemination Layer

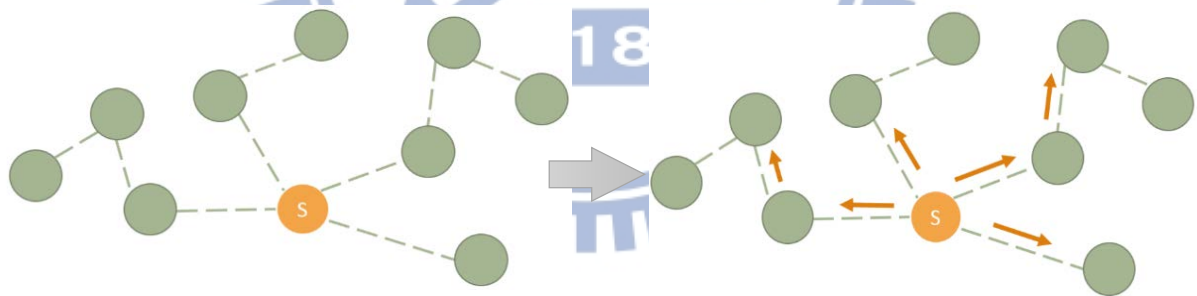


Figure 3: Dissemination of data parameters

As described in section 4.1 of the current chapter, the dissemination layer aims to allow nodes to announce the services available to the network. Since no preceding knowledge about the services available in the network exists in the nodes participating, only the node that contains the accessible service is aware of the same at the beginning. As shown in figure 3, the service in node “S” is available in the network, but only the node itself is aware of such service initially, hence the dissemination process starts taking place to notify the neighbor nodes of the service through a multiple hop method where the message is extended until it reaches a maximum distance allowed by the protocol to avoid unnecessary overhead in the system.

The dissemination layer makes use of the functionalities provided by the neighbor detection and reliable dissemination described in the following sections of the chapter.

4.2.1 Neighbor Detection

The proposed protocol reacts to the changes occurring in the network, with the objective of maintaining the connectivity and available access to the functionalities of the same. During the dissemination phase, the knowledge a particular node has about its neighbors is fundamental to keep the network updated of the available services in the same. The neighbor detection process determines which nodes are located a distance of one jump away from a particular node, all nodes fulfilling the requirement are known as neighbor nodes.

Figure 4 shows the node S and its neighbors A, B, C and D all located one jump away from node S.

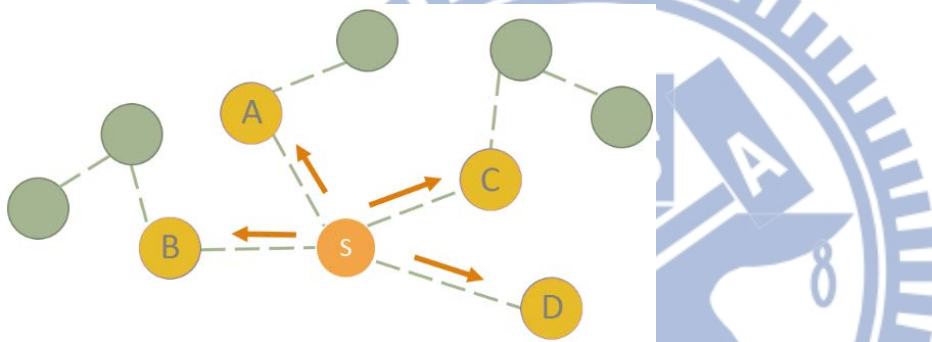


Figure 4: Neighbor detection process

The protocol makes use of the neighbor detection process and maintains the communication by adapting to nodes appearance and disappearance from the network, each node should be aware of the new incoming nodes in the network as well as the ones leaving the network to keep the network updated with the new added information and the ones eliminated.

The protocol makes use of timed beacons (t_B) handled by every node in the network, due to the shared nature of a wireless environment, the transmission of a message will be acknowledged by all the nodes located within a range of the transmitter node, hence these nodes will be considered neighbor nodes fulfilling the one jump distance requirement previously mentioned.

After successfully receiving the message from the transmitter node, the neighbor nodes start forwarding their beacons in a timely manner to keep the alive status in a table created in each transmitter node, if for any reasons no beacon is received in a period of time (T), the status will be changed to unavailable and the information erased from the table.

[39] presents an in depth study of the advantages and disadvantages of different update periods and the considerations of the characteristics of the network at the time of choosing the beacon factors such as reduced amount of outdated information, energy consumption and increased amount of messages in the system. An inverse relation between the mobility of the nodes in the network and the period used for timed beacons can be seen, the greater the mobility of the network, the reduced the period of timed beacons for updates. The period T to decide whether a node is still available in the network is defined by:

$$T \geq 2. t_b \quad (3)$$

The tables contain two entries to determine the status of the neighbor nodes, the node identifier and last occurrence of a message from a particular node. If a new identifier is received it will be added as an entry to the table and the last occurrence or timestamp will be linked to the single identifier. In a real world scenario, the last occurrence of a message only have meaning to the nodes internally, since is not possible that all nodes participating in the network have synchronized clock. When a new message is received but the table already contains the node identifier, the beacon occurrence is updated. Figure 5 show the tables of each node and the different entries created as result of message or exchange or beacon update, in case of node 2, the table contains two entries due to nodes 5 and 8 being in the proximity of node 2.

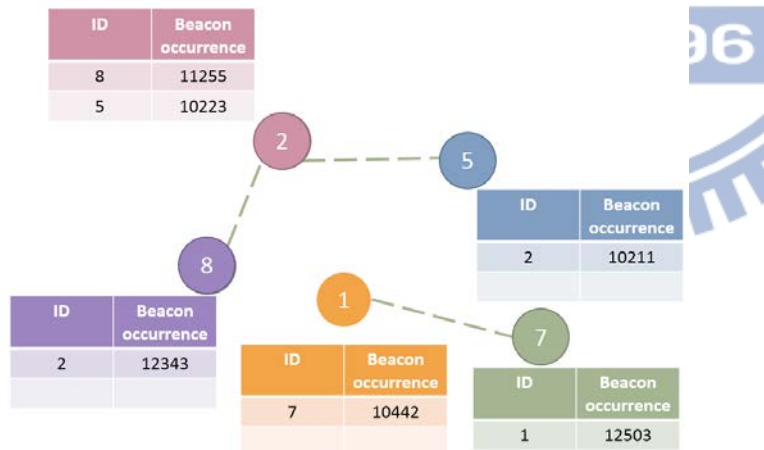


Figure 5: Table entries for neighbor nodes in a network

4.2.2 Reliability

As mentioned in the section 2.1.2, the access to communication is handled by the IEEE 802.11 protocol, but the same cannot guarantee the successful delivery of all messages in the wireless environment due to the different problems (e.g., Interference, reflection/refraction and blocking of signals, etc.)

The present work assumes the existence of a mechanism that guarantees the reliability of message delivery to neighbor nodes participating in the network. When the information is transmitted to nodes in the network, it is assumed such information is known by the receptor nodes. The reliability of the dissemination layer depends on the criteria followed by the nodes receiving the messages, where correct acknowledgement results on sending a response type called ACK (acknowledgement) to the transmitter node of the information. The response contains a single identifier formed by the expression:

$$ACK = \{IDt, count\} \quad (4)$$

Where IDt is the single identifier of the node transmitting the response message and $count$ is a counter that increments with each transmission of a response message.

In section 4.2.1 the tables and entries used by the nodes for the transmission of information between the neighbor nodes was introduced. Using the different entries in tables, a transmitter node creates a waiting list to control the acknowledgement messages received from the different nodes in the network. When an ACK response is received by a node, the identifier is checked and the entry in the waiting list is eliminated, in case a node has disappear from the network, a maximum waiting period (Tw) to receive a response is applied, if the waiting time is larger than Tw a retransmission of the message takes place.

The expression below describes the waiting period, and makes use of the estimated transmission time in the medium (Tt), an estimated time for nodes to process the incoming messages (t) and a random variable ($jitter$) to solve the possible synchronizations among different nodes as described in [40] .

$$Tw = Tt.(targets + 1) + t + jitter \quad (5)$$

4.2.3 Propagation limitation

The type of data handled by each node on the network is a necessary information to build the compatibility between them. The process to perform the comparison and match during a service search is initially built in the dissemination layer. Each node in a network contains a data type table that starts being populated during the neighbor detection process explained in section 4.2.1.

For each node n_i in the network, a table of data type D_i is created, the table is represented as a function $D_i: O \rightarrow H_p$ relating the data type classified through an ontology (O) and the projected total number of hops(H_p) not greater than a maximum number of hops set($Hmax$).

An entry in table D_i is formed by the data type (D_T), projected number of hops (H_p) and single node identifier (nID) that provided the information.

The expression for an entry in table D_i can be written as:

$$D_i \text{Entry} = \{ D_T, (H_p, nID) \mid D_T \in O, 0 < H_p \leq H_{max}, nID \in N \} \quad (6)$$

Figure 6 shows different nodes with their respective tables and the different entries as described previously.

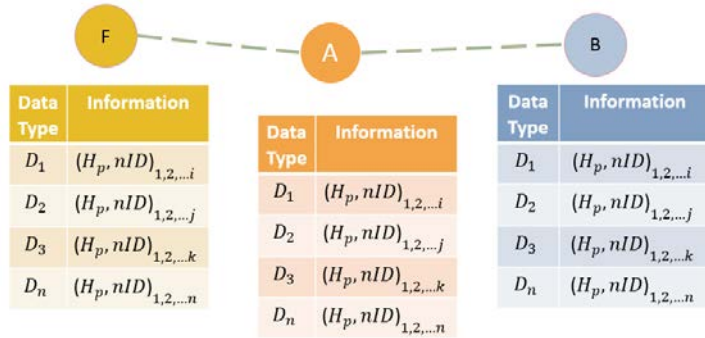


Figure 6: Data type tables for nodes in a network

The tables contain the data type and a series of projection hops and node identifier pairs, to choose the maximum value of the projection hops (H_{max}) a comparison among all the entries' H_p takes place and the highest value is taken as the hop effective value (H_{eff}) to be used for a certain data type D_i . In case a certain node in the network provides two different values of H_p , the most recent information received is the one recorded in the tables, erasing permanently the previous value obtained in a past instance.

The expression below describes the effective value chosen for a certain type of data:

$$H_{eff}(D_i) = \{ (H_p, nID)_{i-1} \mid \text{if } (H_p, nID)_{i-1} \geq (H_p, nID)_i \} \quad (7)$$

4.2.4 Ant colony and Infectious concept for message spread

Ant colony optimization algorithms present a great advantage for resource discovery protocols, different limitations, conditions and characteristics are applied in the concept of these type of algorithms to efficiently spread relevant information to nodes participating in a network while attaining to the limitations and requirements of the same.

The previous section 4.2.3 describes the projected number of hops and the mechanism to choose suitable values for the different types of data handled by the services. The H_p plays an important

role in delimitating the spread of information within a relevant area, minimizing the overhead of the system and decreasing the amount of messages necessary for the dissemination.

Depending on the type of data D_i , carrier ants ant_{D_i} are created using the node identifier nID of the node that started the dissemination process and the type of data D_i to create the base for the route composition process that is described more in detail in section 4.4.

The projected number of hops H_p determines how far the ants of a specific node will move before ending the diffusion process, to avoid multiple carriers following the same trajectory each carrier “infects” the starting point of the trajectory using their identifier $ant_{D_i}i$ to notify other carriers to choose an empty path to follow. The value of H_p for the node initiating the dissemination is equals to $Hmax$ taken from the data type tables as mentioned in section 4.2.3. For each node reached during the trajectory, the value of H_p is decreased by one, this way delimitating the diffusion area and avoid the flooding of the network. If a positive response is found during a trajectory followed by a certain carrier, the route information is brought back to the original node that started the process.

Figure 7 shows a carrier ant_{D_i} and the changes in H_p value through the dissemination process. Starting from node A with data type D_A , a set of carriers ant_{D_A} is created containing the node identifier of node A. For every node reached the H_p value is decreased by one for each $ant_{D_A}i$, when the value reaches one the dissemination stops, hence, the nodes I and J in the figure are not aware of the services offered by node A.

The search process takes advantage of the features of the dissemination layer to facilitate the compatibility match and route composition as will be explained more in details in section 4.3. Node C in the figure represents a response found during the trajectory of a particular ant_{D_A} , the response is carried back to the node that started the search or dissemination process containing the trajectory information to build the route allowing the communication between node A and C.

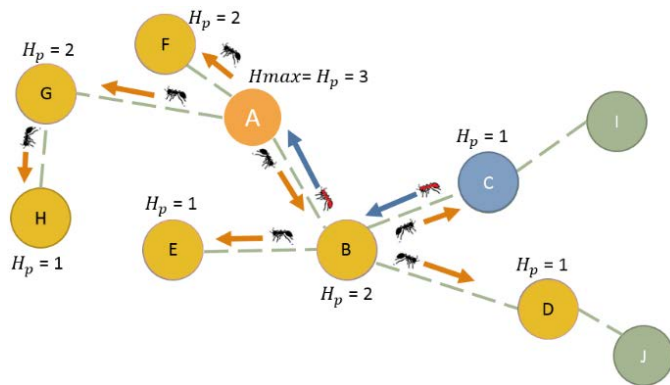


Figure 7: Carrier dissemination and projected number of hops in a network configuration

4.2.5 Network events

One of the characteristics of MANET networks is the mobility of the nodes participating in the environment, the different updates of the information contained in the data type tables rely on the constant notifications and the capability of responding accordingly to the changes in the network or the events in the same. The dynamic nature of the network presents a challenge for service discovery protocols, and the capability of producing proper responses determines the efficiency and competence of the methods. The individual events that take place in mobile networks are presented with the considerations and response of the dissemination layer explained until now.

- Node join

Nodes join the network at any instant and through the dissemination of the data type table the information is passed to the new nodes as well as notifying the neighbor nodes of their current status or services. Due to the shared nature of wireless networks, any message that is broadcast through the medium will be received by nodes located at a certain range of the transmitter. Once the update notifications are sent and the changes in the table are made, the corresponding nodes will notify its neighbors of the necessary changes or updates.

- Node leaving

The same way as nodes joining, a number of nodes may leave the network at any moment. The protocol should eliminate the information related or provided to these nodes that became unavailable or no longer form part of the network altogether. Through a set of processes through update notifications and response time limitations, the entries in data tables are modified to not include obsolete information to later notify the neighbor nodes of the necessary updates in their tables.

- New service availability

In the case where a node wants to make its new service available to other nodes in the network, the data type handled by that particular node are broadcasted through the wireless medium informing of the changes in a previous configuration or the necessary information for a completely new accessible service. The updates in the tables allows the nodes to obtain the necessary information to access the service as well as notify the neighbor nodes of the same.

- Service cancellation

On the other hand, when a node needs to cancel its service for any reason (e.g., power consumption, memory usage, etc.) it notifies the neighbor nodes about the unavailability of the data type previously offered by the node. Through update notifications, the data elimination request is disseminated in the network using the dissemination features and updating the corresponding entries in nodes' tables.

4.2.6 Dissemination algorithm

The node offering a service or updating its status, starts the action by creating a new message variable initially empty that will later contain the service description, or type of data that needs to be added or eliminated from the data table of other nodes in the network.

If there are more than one type in the message, the process will go through the first one and follow up with the next ones until finalizing the queue in the message. As explained in section 4.2.3, the number of jumps and node identifier pairs for a type of data are contained in the Information section of the data table, hence the algorithm access and obtains the necessary information from the relevant column and place it in an array for later comparison and use. If the node is notifying the network of an unavailability of a service, or the waiting time threshold is reached to establish communication with a node, the protocol proceeds to eliminate the relevant information linked to the unavailable service and uses the Carrier function to propagate the changes in the node's table to the neighbor nodes until the number of jumps reaches one. The same way, if a new service appears in the network, instead of the elimination process the protocol performs and add in the table content, with the updated information the carrier function can again propagate the changes to the nodes in the network to update the data tables and have access to the service.

The dissemination layer forms the base in which the search requests are handled as it is explained in section 4.3, the ontology in which the different data types are classified allows the search process to adjust the generic level of requests. Figure 4.2.6 shows the dissemination algorithm and the different variables acting in its functionalities.

```

DisseminationFunction():
  m: message
  Hp: number of jumps
  nID: node identifier
  D: data type table
  n: node
  ant: carrier
  NewMessage m ← Empty
  For each ( data D in m) do
    D.information ← Obtain the set of hop/identifier pairs in data table (Hp, nID)
    if ( ContentEliminated (m, D) ) then
      (Hp, n) ← ObtainDistance (D.information) then
        (Hp, n) ← Eliminate entrance (D.information, nID)
        if Hp > 1 then
          Carrier(m, D, n, Hp)
        end
      end
    end
    if ( ContentAdded (m, D) ) then
      (Hp, n) ← ObtainAdd (m, D) then
        Hp ← ObtainDistance (D.information, D)
        D(i) ← InsertDistance (D.information, Hp, n)
        if Hp > 1 then
          Carrier(m, D, n, Hp)
        end
      end
    end
  end
End
Carrier(m, D, n, Hp)
  Insertinformation (m, D)
  While (Hp > 1) then
    For each ( carrier ant )
      ant(i) = (D, nID, m)
      Hp--
      AddtoTable (m, D, Hp) || Eliminate (m, D, Hp)
      Propagate()
      Check resource availability
    end
  end
end

```

Figure 8: Data type dissemination algorithm

4.2.7 Applications

Service discovery is fundamental for the development of the advanced level of the Internet of Things; automation, interoperability, coupling and others are some of the challenges for current systems. These algorithms will aid in the external components coupling, as well as to efficiently allow services and devices to announce its services to the network and make it available to other

agents in a standardized way, improving the efficiency and security of the system. The system can benefit from the resources in the network, creating a certain level of judgment to define different states, developing predictions of outcomes of different settings and advising on a better solution.

4.3 Compatibility match and search mechanism

The objective of this part of the protocol is to perform the necessary comparisons between data types, decide the level of compatibility of the descriptions of a search and the services available in a network and make use of the ontology classification to find the most suitable response for a specific search performed, allowing for a high level of expression in the search descriptions.

Besides the particular characteristics of the compatibility match and search mechanism, two factors that will be explained in the following subsections are fundamental for the efficient function of the mentioned mechanism.

4.3.1 Service and search descriptions

The functionalities of a particular service can be described in different ways, making reference to the characteristics of the same. The different services in the network should be accessible by other nodes in a remotely manner, including the information, functions and communication method of the same.

The present work makes use of the data type handled by the nodes to create and use the compatibility characteristics between a request and a result. There is a close relation between the type of data used as an input of a process and the type of data that results after the process is performed and finished, the same way, service and search descriptions aim to provide information about the type of data that a certain node can handle or is looking for, respectively.

Although a number of technologies providing services had been developed throughout the years (e.g., REST, Web Services, etc.), which solution is chosen for the proposed protocol remains irrelevant, since the mechanism assumes the existence of a technology to perform the comparison of data type of a service. The classification architecture can be compared to the concept of semantic web where services are described semantically using the concepts of the same with the advantage of being capable to be processed automatically and determine the relation between the concepts used in the descriptions of [41] . The same way, using the ontology classification, explained in more details in section 4.3.2, the relation in the description of a service includes the type of data handled allowing relation building capability.

In the previous sections a different set of variables was mentioned such as the node identifier that contains the single identifier of a node and is used in the dissemination process to provide the information of which node started the action, besides that, there are other variables that need to be included in the description for the mechanism to perform correctly and efficiently:

- Service identifier

Refers to the identifier of the node offering the service in the network and through which the other nodes in the network obtain the information about the query starting point. The service identifier is fundamental in the creation of a subset of identifiers that are used in different purposes such as the unique search identifier that is explained in more details in section 4.3.3 as well as the ant_{D_i} for the propagation method as described in section 4.2.4.

- Action

As in any process, the type of data a node is able to handle does not provide an accurate description of the actions or functions the node can perform with the data received or produced. Hence, including the action narrative capable to be executed complements the service and search description characteristics.

- Data type

Classified through an ontology entirely diffused in all nodes that participate in the network, a service can define a value n of data type that is later used in the compatibility match through best effort comparison. As described previously, a process is formed by the input data, the function performed and the result obtained after finalizing the process. The same way, services define two types of data InD_i and $OutD_i$ referring to input and output type of data D respectively.

4.3.2 Ontology Classification

Refers to a list containing the different concepts that can be related to the types of data used in the description of services for the nodes in the network. The objective of creating such classification is to build the relation between the different specifications providing a high level of expression for search processes of services. As described in chapter 3, different solutions had taken advantage of creating service classification to be used by the different protocols. Including the ontology features to the classification allows for a new approach that makes use of the advantage of the relations to enhance efficiency. Different levels of relations can be

created between the concepts using dedicated ontology languages such as OWL, although for the present work, only the relation between data type is applied, more advanced relations can be considered for a future improvement of the mechanism. The OWL Web Ontology Language Overview [42] presents all the functionalities, details and capabilities of the mentioned ontology language as well as the different levels of relations between the concepts included in the same.

Figure 9 shows a representation of an ontology containing different concepts, the relations between the concepts are shown by the circles in the figure, and these relations at the same time can be configured as one-to-many or many-to-one:

- Identical

When two sets present equal characteristics, hence is assumed they have the same type of data.

- Inclusion

When a certain type of data is more general than a second type. In the example figure, type A includes B which at the same time includes C, hence is safe to assume the information in A is more generic and includes the ones described in B and C.

- Unrelated

The types do not share any common information expressed by the ontology, as it would be the case of A and D in the figure.

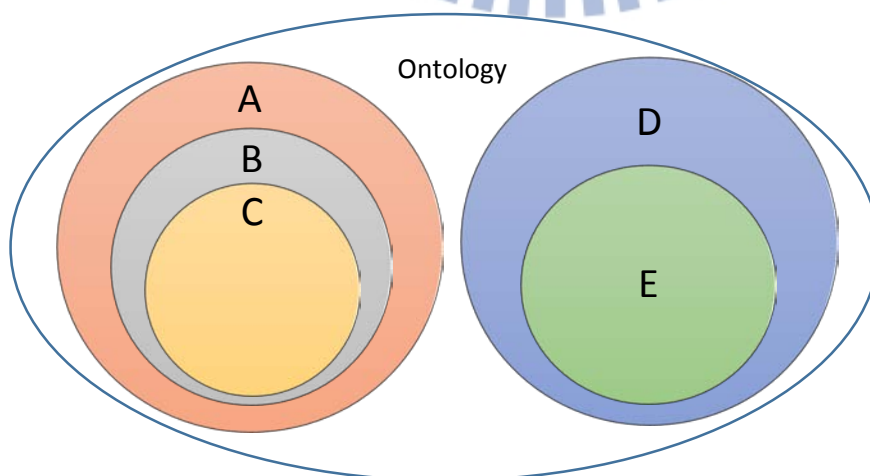


Figure 9: Representation of relations between concepts in an ontology

As mentioned in section 3.2.2, the present work assumes the nodes participating in the network have a predetermined knowledge about the ontology used for the classification. A proposed solution to overcome this limitation would be a dissemination phase of the ontology to joining nodes allowing them to operate through the protocol, but a deeper consideration of the challenges, such as the presence of multiple ontologies in the network and correct selection of suitable one for protocol compatibility, as well as the feasibility of this proposal, is necessary. Placing this and other considerations as an individual research and development topics for future investigations.

4.3.3 Search process

The objective of this part of the protocol is to find a particular service that fulfills the requirements of a search request. The dissemination layer makes use of the data types classified in an ontology used by all the nodes in the network, the compatibility match and search mechanism takes advantage of the features of data type dissemination to perform the search of services in the ad hoc network, more details about data type dissemination can be found in section 4.2 of the current chapter.

To initiate the search function, a service description containing the characteristics of data type handled by the particular node is necessary as explained in section 4.3.1. When a node is interested in finding a service, a search request is created by the particular node containing the variables describing the service to be found in the network. The expression of the request search message is described by:

$$R_S = \{ID_R, (D_{S1}, D_{S2}, \dots, D_{Sj})\} \quad (8)$$

Where ID_R is the unique search identifier created using the node identifier nID of the node that created the search request, and a local counter handled by the node to differentiate different search requests created simultaneously by a particular node. The variable D_{Sj} refers to the description of the particular requests to be fulfilled by a service in the network, a single or multiple descriptions can be contained in a single search request. Each description uses the input and output data type sets that describes the services and perform the compatibility match as described in section 4.3.1, along with a Time-to-live (TTL) value to control the search request propagation area. The expressions describe the search identifier and service description respectively:

$$ID_R = (nID, count) \quad (9)$$

$$D_{sj} = \{(InD_{j_1}, InD_{j_2}, \dots, InD_{j_k}), (OutD_{j_1}, OutD_{j_2}, \dots, OutD_{j_k}), TTL_j\} \quad (10)$$

Where, $O \rightarrow \text{Ontology}$ and $InD_{j_k} \in O, OutD_{j_k} \in O, 0 \leq k$

The use of the ontology allows for a classification of the data table parameters and facilitates the comparison of the InD_i and $OutD_i$ values of the data table entries of the search request and the ones contained in the services of nodes participating in the compatibility match process. Section 4.3.2 details the different relations between possible during the comparison, besides, figure 9 shows the hierarchy structure followed by the ontology and how different parameters can be contained into one another.

This characteristic allows the compatibility match process and search mechanism to perform in a general or specific manner:

-Specific search

The compatibility match is performed only if the condition specified in the description is matched exactly as described, subsets and similarities are not taken in consideration and will not be included in the search results.

-General search

The search is performed throughout the network with results containing subsets and results with similarities to the description used in the process. With a more relaxed search condition, the amount of results matching the criteria is higher, hence although a certain degree of freedom is allowed, there must still be a consideration during the process.

To improve the efficiency of the search process, a selective mechanism can be applied to function after the results are found, and uses different criteria to choose the most suitable service to complete the request, such as power consumption, speed, location, etc. The current work does not make use of such mechanism for the many considerations and challenges that imply, hence is proposed as a future improvement of the present work and future research topic for related investigations.

When a match is found in a certain node, the same has to answer with a response message directed to the node that initiated the search process. The message contains the description of the service that was found as a match for the request, as well as the node and service identifier of the node offering the service in the network.

The request response message is described by the expression:

$$R_R = \{ID_{Service}, (D_{S1}, D_{S2}, \dots, D_{Sk}), nID \} \quad (11)$$

Being $ID_{Service}$ the identifier of a particular service in the network, D_{Sk} the set of service description satisfying the request and nID the single node identifier of the node containing the service. The service description is formed by the set of input and output data type that fulfill the search requirements, expressed as:

$$D_{Sk} = \{(InD_{k1}, InD_{k2}, \dots, InD_{km}), (OutD_{k1}, OutD_{k2}, \dots, OutD_{km})\} \quad (12)$$

Where, $O \rightarrow Ontology$ and $InD_{km} \in O, OutD_{km} \in O, 0 \leq m$

The information contained in the request response message will be used by the route composition system to create to communication link between the node looking for a service and the one providing it, as it is explained in section 4.4.

4.3.4 Network change adaptability

The mobility of nodes in an ad hoc network represent a challenge for the compatibility match and search mechanisms since not only the movement of nodes causes the topology to change, but also the network is susceptible to nodes arriving and leaving the network at any instant.

To deal with the changes in the network, when nodes receive a search request, they create and add an active search list containing all the searches currently running in the network. The information contained in the list is used to notify new nodes joining the network using a dissemination of the relevant information as described in section 4.2.

On the other hand, when a node leaves the network, the information contained in that particular node should be eliminated from the list and data table of other nodes. The communication routes are directly affected, and a recovery mechanism has to take action to ensure the connectivity among nodes, more details are described in section 4.4 of the chapter.

The active searches in the list remain alive until a response is received from a node in the network or a cancellation request is sent by the node that initially started the search process. In case a node decides to apply modifications to its active search request description, the cancellation request can be used by specifying explicitly the descriptions to be eliminated from a search. Hence, the cancellation request contains the search identifier ID_R and the set of service

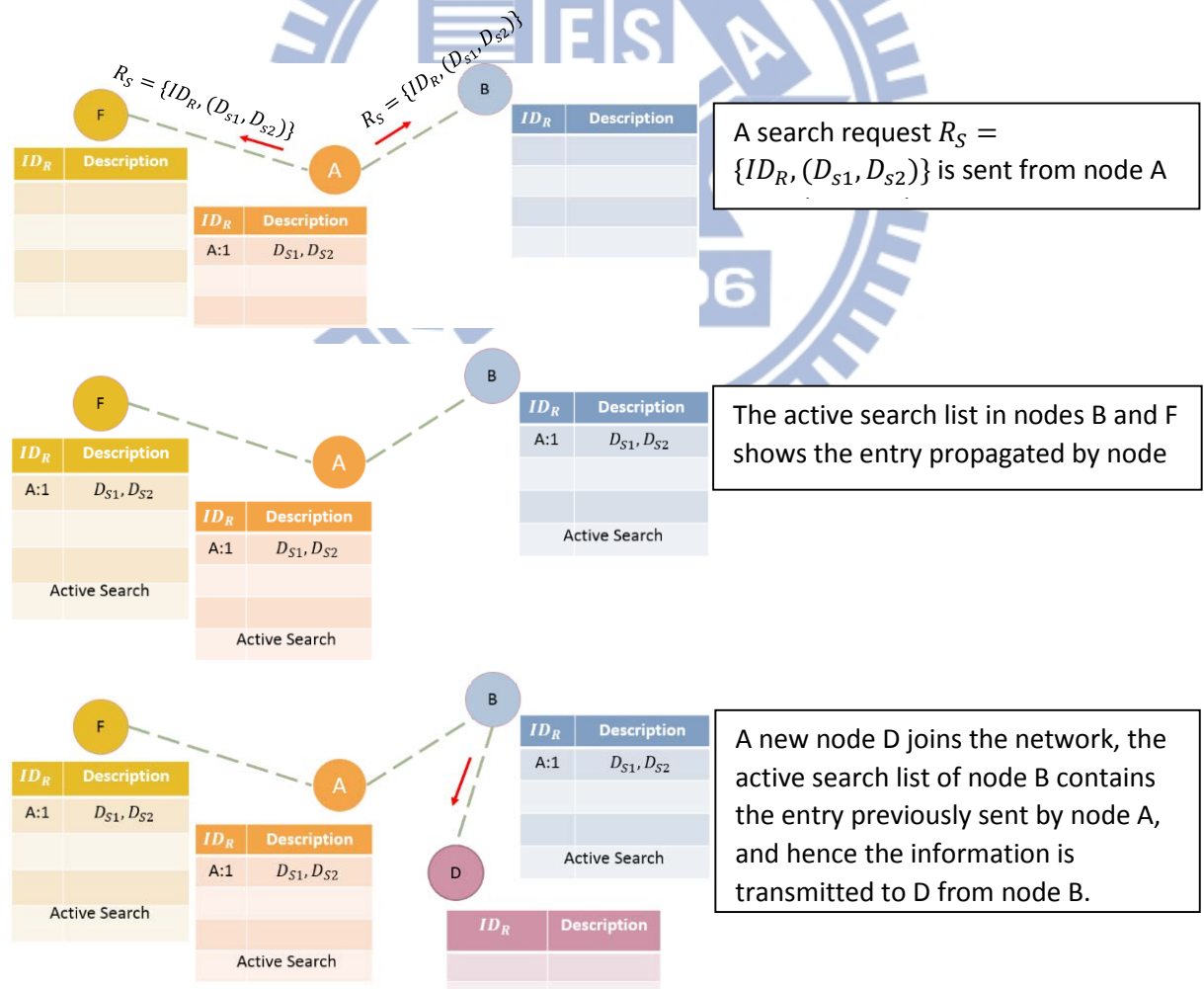
descriptions to be modified or complete cancel command *abort* as the second parameter in the expression respectively:

$$R_C = \{ID_R, (D_{S1}, D_{S2}, \dots, D_{Sj})\} \quad (13)$$

$$R_C = \{ID_R, abort\} \quad (14)$$

The nodes in the network that receive the cancellation request perform a check on the current entries of their active search lists and perform the modifications or completed cancellation of the search, disseminating the request to the neighbor nodes until reaching the maximum number of hops for message spread. If a node that initiated a search requests is no longer available in the network, the neighbor nodes will be notified through the neighbor detection procedure explained in section 4.2.1 and cancel the active search from the list.

Figure 10 shows the adaptability to network change for a node joining the network and cancellation of an active search among nodes participating in the same.



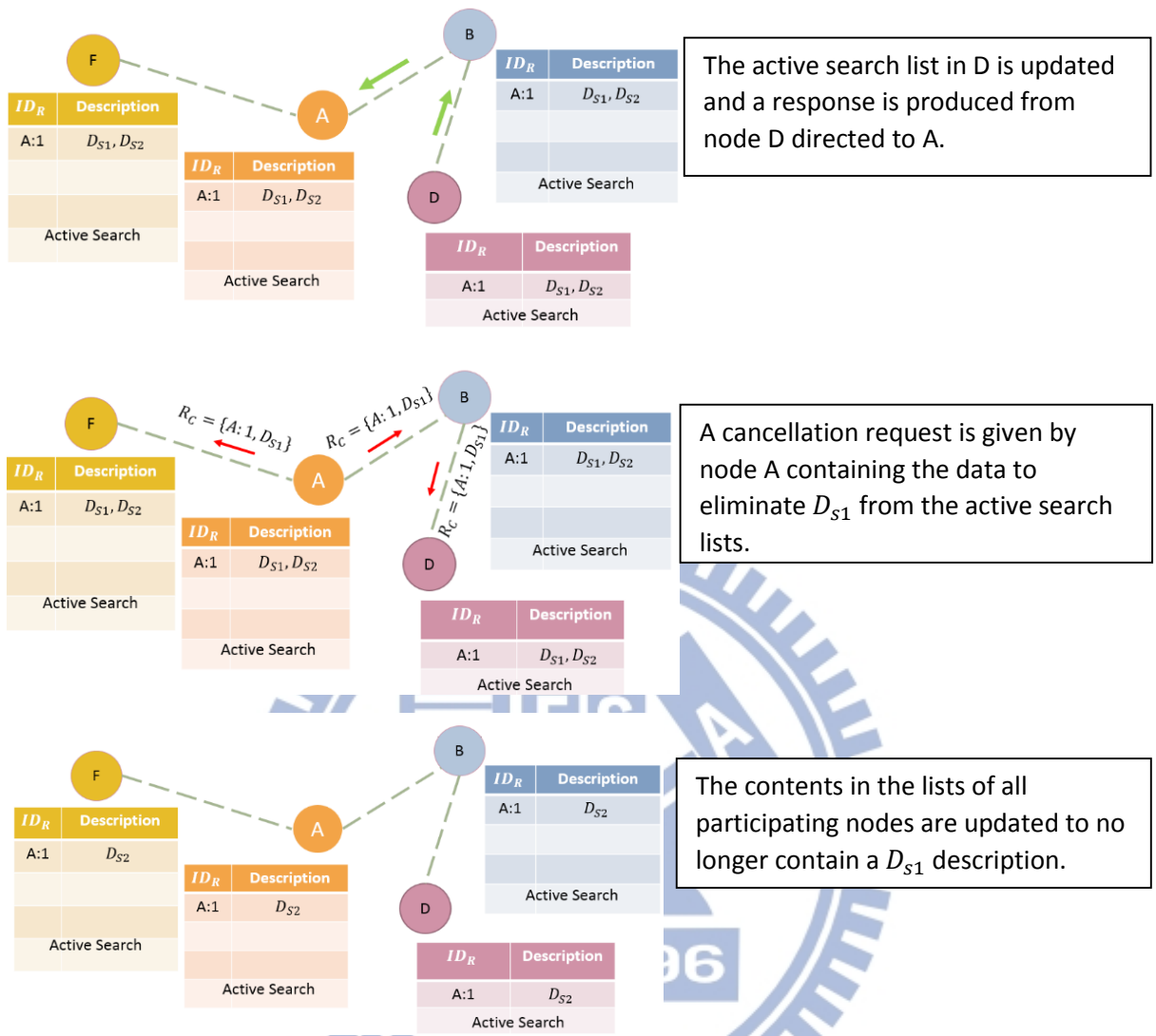


Figure 10: Search mechanism adaptability to network change and cancellation request process

4.4 Route Composition

The dynamic nature of mobile ad hoc networks requires solutions capable of adapting to the constant changes and variations among nodes in the network. The route composition has the objective of constructing the communication link between the nodes initiating a search process and the different services that fulfill the requirements of such search description. As opposed to fixed topology networks, the routes descriptions and construction cannot be assumed to be part of nodes' knowledge, hence, mechanisms that create the links and allows nodes to communicate with each other are an essential part of a service discovery algorithm.

Section 4.2 mentions the necessary features used in the dissemination process that allows to trigger the route composition function, the following sections present the details of the same and operation mechanism to create the connectivity links.

4.4.1 Connectivity bridge formation

To create the connections between two particular nodes, a couple of considerations have to be taken, the first one is the group of neighbor nodes that can transfer the message N_{links} and the second one is the group of nodes that can be reached N_{target} . The group of neighbor nodes N_{links} contains the node identifier information of the nodes participating during the message transferring.

In the dissemination layer explained in section 4.2 refers to the use of the data type tables containing the type of data handled by the node as well as the projection hops and node identifier of the service. The connectivity bridge formation works on the same principle of tables, each node in the network contains a Link Table (LT) used to record the known paths to different nodes that are relevant to others. The table relates the two groups mentioned previously as follow:

$$LT: N_{target} \rightarrow N_{links} \quad (15)$$

For a particular node i , the link table entry in its particular LT_i is called a bridge B . Multiple bridges can be contained in a single LT_i . A bridge expression is described by:

$$B_k = \{ID_B, n_{target}, n_{link}\} \quad (16)$$

Where ID_B , is the single identifier of a particular bridge, n_{target} is the destination node reached through the bridge and n_{link} is the intermediate node to transfer the message towards the final destination. When a request is initiated in a node, the set of carriers ant_{D_i} use the information contained in the link tables to transfer the message towards a particular node, as well as providing the information to create new entries if a response message is created from a node.

Figure 11 shows a graphical representation of the link tables of different nodes and the bridge created during message transfer.

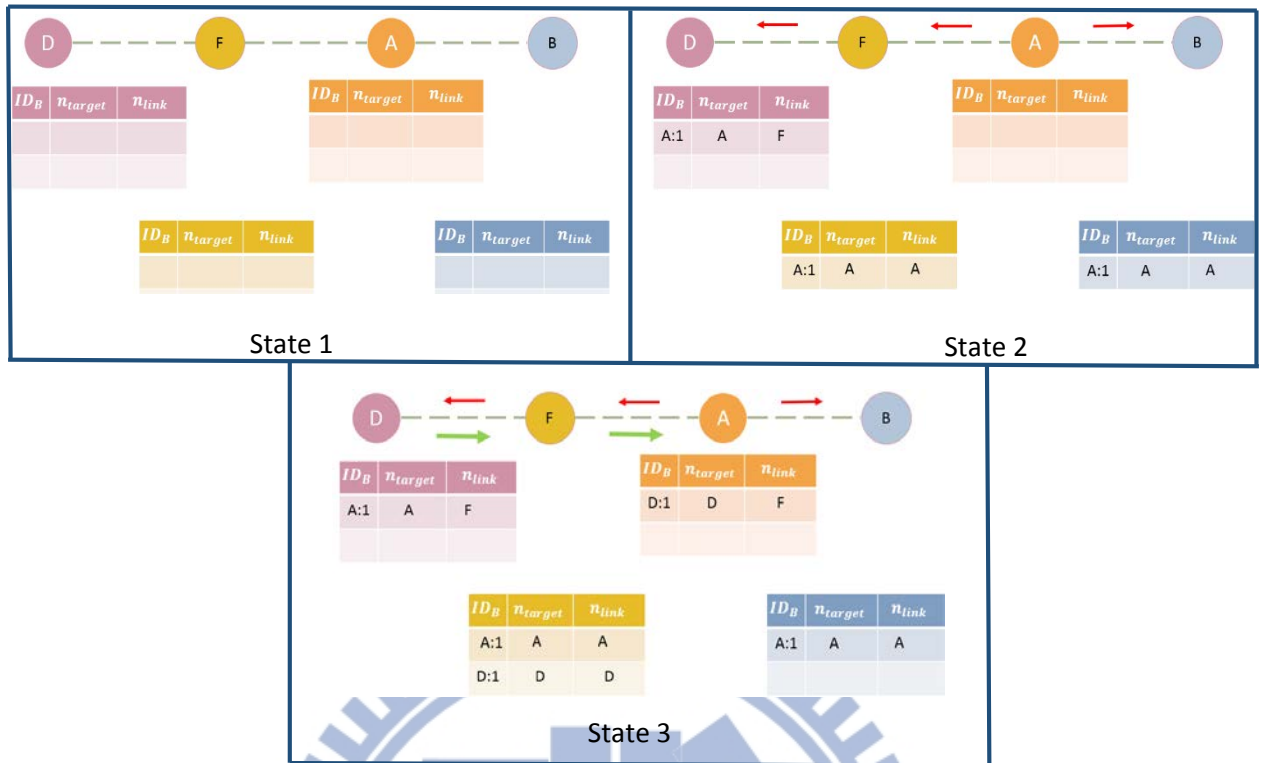


Figure 11: Link table and bridge formation in mobile ad hoc configuration

State 1 shows the initial state before a message is transferred in the network, the link tables LT are show for each node in the network with the $ID_B, n_{target}, n_{link}$ fields for each bridge entry.

State 2 shows a request initiated in node A and disseminated to the neighbor nodes B and F, the bridge entry is created in the link tables of the nodes receiving the message, the bridge identifier ID_B is created using the node identifier nID and a message counter exclusive to the link table of a particular node in the network. The fields n_{target} and n_{link} contain the final destination node and intermediate node through which the message will be transferred, respectively.

State 3 shows a response message generated from node D directed to the request started by node A, hence, the entries in the link tables of node F and A contain the bridge identifier ID_B of node D and the information of n_{target} and n_{link} is transferred to the entries of the nodes, creating the bridge that allows the communication within node A and node D. In this case, the link table in node B does not contain any entry to connect to node F or D, hence the communication between these particular nodes is not possible.

The communication paths or bridges are defined by the bridge identifier ID_B , two different paths cannot have the same identifier. Hence, the process can create multiple paths that connect two particular nodes, but a single path cannot share its identifier with another. In case of nodes

joining or disappearing from the network, the connections created in a previous instance are constantly updated through the neighbor detection process explained in section 4.2.1.

As nodes join or leave the network, the connections should be modified to maintain the most optimal result in the tables. Different characteristics can be considered to choose the most ideal communication bridge to be recorded in a table, for the current work, the shortest path method is applied in the connectivity bridge formation process to achieve the result. Many factors and considerations can be studied for the development of more accurate or improved performance mechanisms, being a complementary part of the present work, it remains as a proposed improvement or future research topic for investigations.

4.4.2 Route control

Section 4.3.2 explains the search process that initiates when a node starts a search request, the message is propagated to the neighbor nodes. The route composition takes place after the search process is initiated, creating the bridge entries with the responses received from the different services available in the network that fit the search description requirements.

To adapt to the mobility of nodes in an ad hoc network and validate the entries in the link tables of different nodes in the network, update messages containing the necessary modifications or eliminations of entries in tables is applied. If a neighbor node n_{link} leaves the network, the destination nodes n_{target} originally reached through the mentioned n_{link} are no longer available, hence the entries in the link tables containing the node no longer available should be modified or eliminated from the tables.

The elimination messages follow the expression:

$$m_{elim} = \{nID, ID_{B_1}, ID_{B_2}, \dots, ID_{B_j}\} \mid j > 0 \quad (17)$$

Where nID is the single identifier of the node sending the m_{elim} message, and ID_{B_j} is the multiple bridge identifiers no longer available and that should be eliminated from the link tables of the different nodes participating in the communication. The dissemination of the message takes place through multiple hops among nodes in the network and each performs a check in their table entries and eliminate the matching results. The multiple paths that lead to a single destination node are only affected if it contain the bridge identifier specified in the elimination message, else the destination can still be reached and the communication link between the participating nodes is kept alive.

Figure 12 shows the elimination message function when a connection between two nodes is lost and the modifications in the link table entries of the affected nodes.

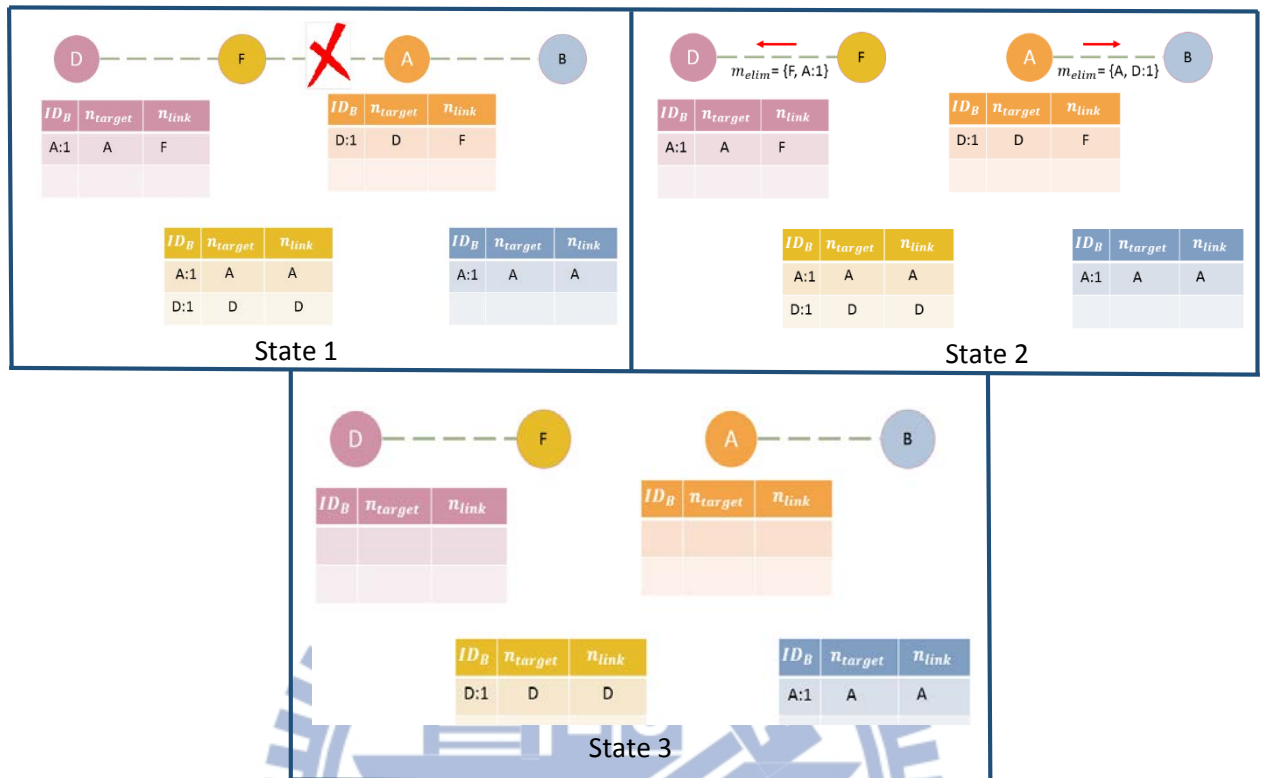


Figure 12: Bridge elimination after lost connectivity between two nodes in the network

State 1 shows the entries in the link tables of the different nodes in the network when the connection between node A and node F is broken.

State 2, the elimination message m_{elim} containing the node identifier of the nodes sending the message and the bridge identifier ID_B to be eliminated from the link tables of the nodes containing a match for the same. Node F produces a m_{elim} containing $ID_B = A:1$ which is the bridge identifier that was used to connect node F with node A, on the other hand, node A produces an elimination message with D:1 as bridge identifier.

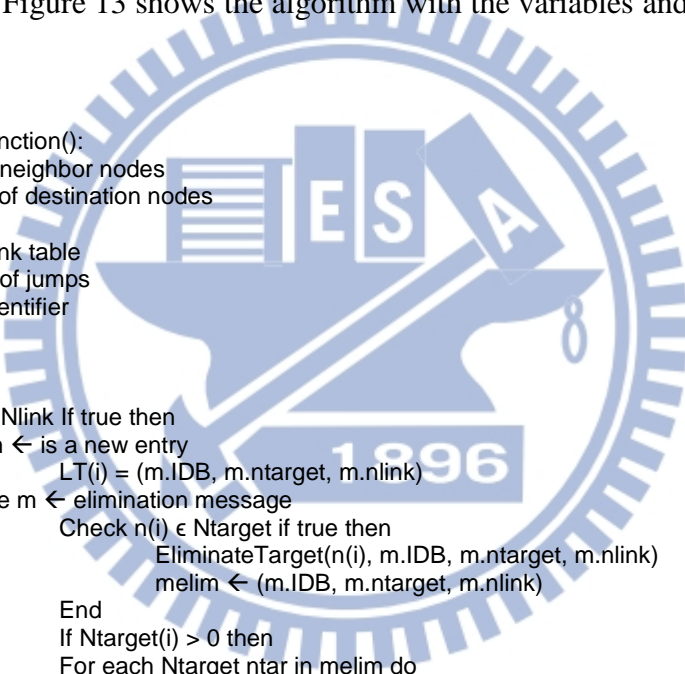
State 3 shows the modifications in the link tables occurred after the elimination message is disseminated to the nodes in the network and the same perform the elimination process if a match is found in their table entries.

4.4.3 Route composition algorithm

The route composition function reacts to the events created in the network by nodes requests. When a message is received by a node it performs a check to handle the message as a new link table entry in the particular node or an entry elimination due to node disappearance or link

connectivity loss. The bridge entry is created in the link table LT containing the bridge identifier, the intermediate communication node to carry the message and the destination node represented by the expressions ID_B, n_{link} and n_{target} , respectively. As mentioned in section 4.4.2, the elimination message contains the node identifier and the bridge identifiers of the entries that need to be eliminated from the link tables of participating nodes.

When an elimination message is received, a check of the different destination nodes is handled by the algorithm, the data received is stored in a variable to later disseminate through the neighbor nodes before being eliminated from the table of the respective node. The *call_carrier* sub-function is activated to start the dissemination of the message following the method explained in section 4.2 and this way allowing the neighbor nodes to make the proper changes in their link tables. Figure 13 shows the algorithm with the variables and data operating in the same.



```

RouteCompositionFunction():
  Nlink: set of neighbor nodes
  Ntarget: set of destination nodes
  m: message
  LT: node's link table
  Hp: number of jumps
  nID: node identifier
  n: node
  ant: carrier
  For each m do
    Check n(i) ∈ Nlink If true then
      If m ← is a new entry
        LT(i) = (m.IDB, m.ntarget, m.nlink)
      Else m ← elimination message
        Check n(i) ∈ Ntarget if true then
          EliminateTarget(n(i), m.IDB, m.ntarget, m.nlink)
          melim ← (m.IDB, m.ntarget, m.nlink)
        End
        If Ntarget(i) > 0 then
          For each Ntarget ntar in melim do
            GetNeighbor (ntar, melim data)
            call_carrier(melim)
          end
        end
      end
    end
  end
end

call_carrier(m)
  While (Hp > 1) then
    For each ( carrier ant )
      ant(i) = (nID, LT, m)
      Hp--
      Eliminate (m, LT)
      Propagate()
      Check resource availability
    end
  end
end

```

Figure 13: Route composition algorithm

V. Protocol Assessment

The network simulator chosen for assessment of the proposed service discovery protocol is the discrete event simulator ns-2, popular and widely used in fields targeted at networking research for its support of TCP, routing and multicast protocols in both wired and wireless networks. Section 2.2 expands the considerations and reasons of using network simulator results as validation for a protocol or solution, also a comparison and more detail study of the different network simulators, the capabilities and limitations of the same, can be found and referenced in Table 1 in chapter 2 of the present work.

The protocol is implemented using the Java Development Kit 7 or Java SE 7 with the update u65 as recommended by Oracle for its latest security fixes. The standard communication libraries and packages includes networking functionalities as well as sockets communication using either Datagram sockets or UDP sockets for transport of messages. The current work makes use of UDP sockets as the main communication mechanism, transferring messages and operating in the dissemination of the same between nodes in the ad hoc network environment.

The ns-2 network simulator allows to run a Java protocol using the AgentJ extension, hence the mentioned code written in Java is able to run in every node of the network in the simulation environment. The work in [43] presents a modification of the AgentJ extension and explains more in detail the functionality and capability of the same to allow Java protocols to run within ns-2 network simulator.

The network simulator requires a set of parameters to be set for operating in adequate conditions, the Table 2 presented in chapter 2 define the wireless communication and other aspects for the simulation, besides the mentioned parameters, section 5.1 introduces the settings and specifications that influence the outcome of results in the testing period.

5.1 Network and environment configuration

The environment for the simulation is created using three machines forming a secured local area network where nodes are free to move within the same. The environment is controlled in the main machine handling all the initial settings and triggering the random transmission of nodes throughout the network. Ubuntu operative system is installed and used in all machine of the network for the easy connectivity, robust behavior and fast accessibility to cloud server for data record.

To simulate the mobility of nodes in the network, the random way-point model is chosen for its characteristics, different mobility models are presented in the work [22] . The model allows nodes in the network to move using a range of velocities specified in the simulation parameters and random directions located within the provided network space and simulation area, an *idle* period is applied so nodes will remain in a certain position for a period of time before moving to a different location.

Figure 14 show the configuration of the environment where the simulation takes place.

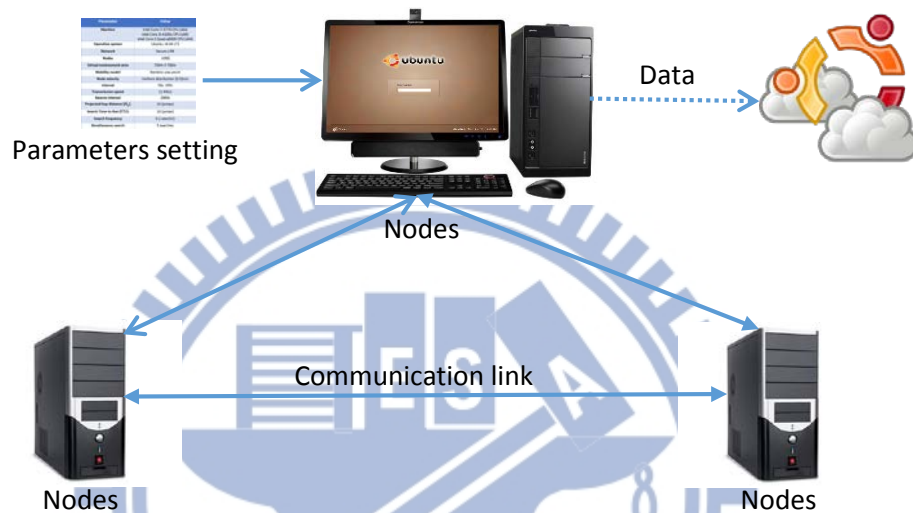


Figure 14: Simulation environment for service discovery protocol

The parameters are set in a main machine controlling the settings and specifications to follow during the simulation, Table 3 shows the different parameters and the values selected for each for the execution of tests. A series of twenty repetitions have been made for each section of the tests to obtain an average of the results represented in the graphs, each execution generates data that is passed to a cloud server that handles the comparison and calculates the average for the final result to be displayed for a specific section of the simulation.

Table 3: Service discovery protocol parameters configuration for simulation

Parameter	Value
Machine	Intel Core i7-3770 CPU (x64) Intel Core i5-4200u CPU (x64) Intel Core 2 Quad-q6600 CPU (x64)
Operative system	Ubuntu 14.04 LTS
Network	Secure LAN
Repetitions	20
Nodes	1000
Virtual environment area	800m X 800m
Mobility model	Random way-point
Node velocity	Uniform distribution [0.5]m/s
Mobility Interval	100s, 200s
Transmission speed	11 Mb/s
Beacon interval	2000s
Maximum projected hop distance (H_{max})	10 (jumps)
Search Time-to-live (TTL)	10 (jumps)
Search frequency	0.2 search/s
Simultaneous search	15 searches

Deciding an appropriate simulation scenario is a complicated but important process, the work [20] presents several models that take into account two metrics for the creation of simulation scenarios that meet the evaluation standards:

- Average shortest-path hop count

The transition of a message from a node to a neighbor node via a communication link is defined as hop. The average shortest-path is the shortest average distance that can exist between two nodes in the network. In the mentioned work the average shortest-path hop count is taken by the equation:

$$A_{sp}Hops = \frac{\sum_{t=1}^T hops_t}{\sum_{t=1}^T paths_t} \quad (18)$$

Where T is the number of snapshots constructed, $hops_t$ is the total number of hops among all node pairs at time t and $paths_t$ is the number of non-zero elements at time t . For more details refer to the bibliography in [20] .

- Average amount of network partitioning

Refers to the proportion of node pairs that do not have a communication path between each other. The ANP (average network partitioning) is defined by the equation:

$$ANP = \frac{\sum_{t=1}^T z_t}{n(n-1)T} \quad (19)$$

Where z_t is the amount of node pairs without a connection path in between at time t , n is the number of nodes making $n(n - 1)$ the potential number of links and T is the number of snapshots. Low network partitioning increase the possibility of message delivery during the evaluation of a protocol in MANET networks.

Following the equations presented above with the characteristics of the environment used in the present work the results obtained for average shortest-path hop count and average amount of network partitioning respectively are:

$$* A_{sp}Hops = 4.15$$

$$* 0 < ANP \leq 5\%$$

Hence, the values for H_p and TTL in Table 3 are set much higher than the $A_{sp}Hops$ required, allowing the messages to spread and fully cover the whole network used in the simulation, increasing the possibility of fully discovering the services scattered across the network and testing the protocol in an adequate scenario.

5.2 Dissemination layer message reduction

The dissemination relies on the transfer of messages hop by hop between nodes participating in the network, as described in section 4.2, a service is originally known by the node that offers it in the beginning, for other nodes to be aware of the available service, the dissemination takes place using messages to notify other nodes of the services. Similarly, the compatibility match and search mechanism makes use of the features of the dissemination layer to spread the search descriptions and find services that meet the requirements of it.

This particular part of the testing phase aims to compare and show the reduction of messages use during the dissemination layer and other processes relying on message spread when the data compatibility service discovery algorithm runs.

The graphs show a side by side comparison of the results when the algorithm does not run and when it does. As specified in the simulation parameters, a total of one thousand nodes is randomly spread across the virtual networked environment with the liberty of moving freely across the network. A variable amount of services from 20 to 200 are disseminated in random nodes in the environment with each service containing six data type (input & output) classified by an ontology.

Figure 15 shows the results of the comparison when the compatibility match service discovery is running against the outcome of a traditional service discovery protocol that does not make use of any compatibility mechanism among data types handled by the nodes. Two different node mobility intervals are applied during the simulation, where at first nodes will move every 100 seconds and the second set is the result when nodes move every 200 seconds.

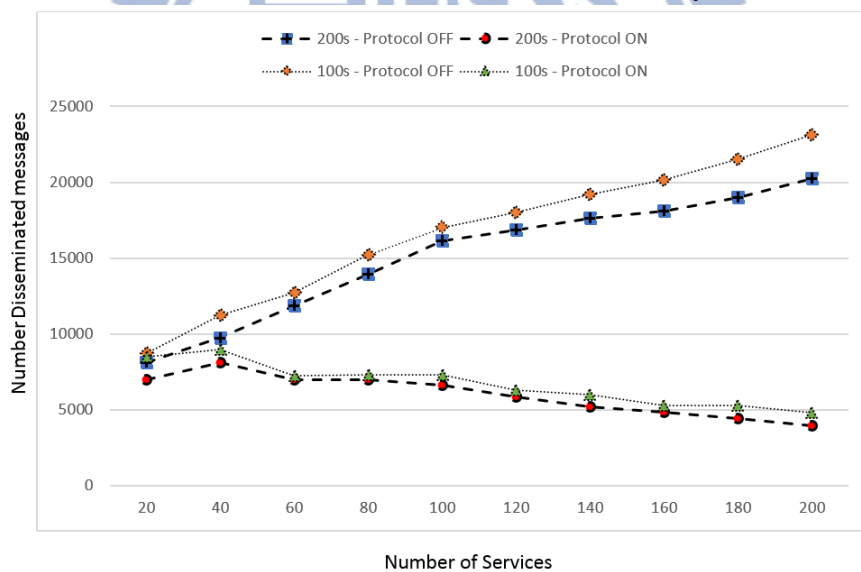


Figure 15: Reduction of the amount of transferred messages in the dissemination process

Figure 16 shows the performance of the compatibility match and search mechanism when using the data compatibility service discovery protocol against a traditional protocol with no data type match capability. Every search request started by a node in the virtual environment is kept active for 10 seconds, after the time limit is reached, a cancellation request R_C , as mentioned in section 4.3.4, is sent by the same node, the mobility intervals remain the same as previous. A total of 15 nodes participating in the network are selected randomly every 5 seconds to start the search requests simultaneously with a total running time of 100 seconds.

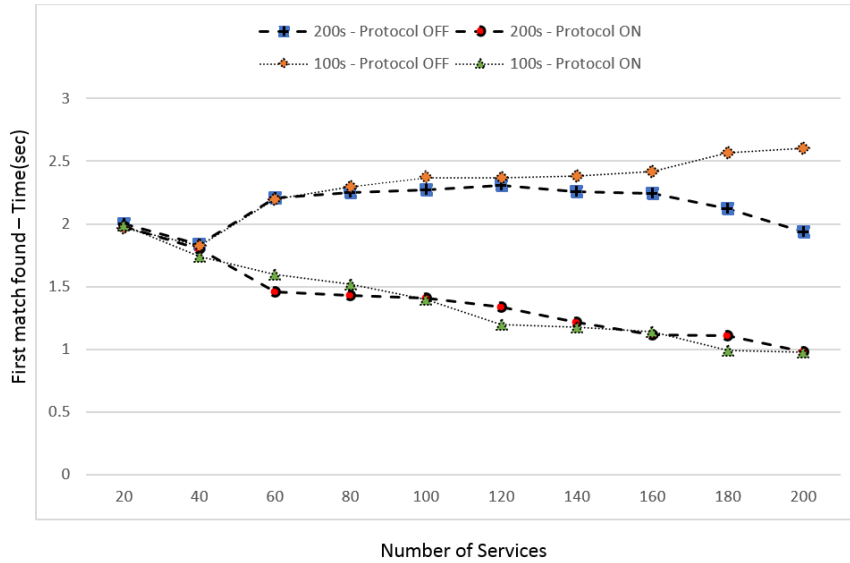


Figure 16: Average time for request/service compatibility match in search process

The results shown in the graphs above demonstrates the influence data type compatibility has during the dissemination process by reducing considerably the amount of transmitted messages. Also, by using the data type classified in an ontology, the relation allows for greater expression during the search process, decreasing the amount of time to find a service compatible with a search request description. When the amount of services in the environment increases as shown by the X axis of the graphs, the search time decreases because there are more available services that fit the requirements of a particular search. Without the data type classification, the number of messages necessary in the dissemination increases directly with the amount of services in the network, the data type classification through the ontology aids in the process by creating blocks of data related services improving the performance of the system by decreasing the amount of necessary transmitted messages.

Section 3.2 of chapter 3 studies the current state of service discovery algorithms and presents different solutions classified in different categories. To compare the performance of the proposed solution with the works mentioned previously, the parameters used during the simulation and data gathering should be the same as the ones used in the present work. Unfortunately, the main problem faced by researchers when comparing solutions is the lack of specification for the parameters used in the different simulations, this translates to the impossibility of accurately comparing the results obtained with the results published in previous works related to mobile ad hoc networks. The section 2.2.2 presents the importance of parameter settings and the heavy influence any minor change has in the outcome results. An

accurate comparison between the solutions will only be correct when using the exact same parameters to produce the results of different solutions.

An exact specification of the functionalities and logic of the different solutions is not available in the published works, hence, a near similar implementation of the different solutions will not provide an accurate comparison without recurring to the original codes designed by the different authors. For these reasons, the present work shows a subjective comparison of different approaches that aim to solve an identical problem.

5.3 Protocol reliability through miss hit occurrence

The miss hit effect occurs when the system finds a positive match for a particular request description but the data type parameters that satisfy the request does not exist in the network. If a protocol's performance oscillates greatly in the results obtained, the protocol is considered faulty.

To simulate the behavior in the present work, 300 nodes in the network were randomly selected and assigned a total of 6 data type parameters (input & output) for each node to be used in the compatibility match and perform the service discovery task. A variable named FRR (False Response Ratio) controls the ratio of invalid search results produced during the active service discovery phase of the protocol. The higher the ratio, the more difficult it is to correctly discover the services in the network.

Figure 17 shows the performance of the service discovery protocol when the FRR value is increased and how the system behaves to the characteristics of the network for two different sets of nodes with mobility intervals of 200 seconds and 100 seconds. As specified in the section 5.1 of the current chapter, the configurations of the environment gives an *ANP* value greater or equal to 5% following the equation presented in [20] .

This implies that at any given moment, a total of close to 5% of nodes are disconnected from the network and hence, the percentage of services discovered in the network cannot reach the total of 100%. The reasons for this is that the connectivity paths between a node starting a request and a node providing a suitable service do not exist or it was broken, also, the TTL value of a request might not be enough to reach a service that fulfills the requirements of the same, without the chance to reach all the nodes in the network, the service discovery is reduced, another reason is the possibility that a node that started the search or one that offered the service

suddenly changed locations or left the network, resulting in the invalid search request message and request response communication.

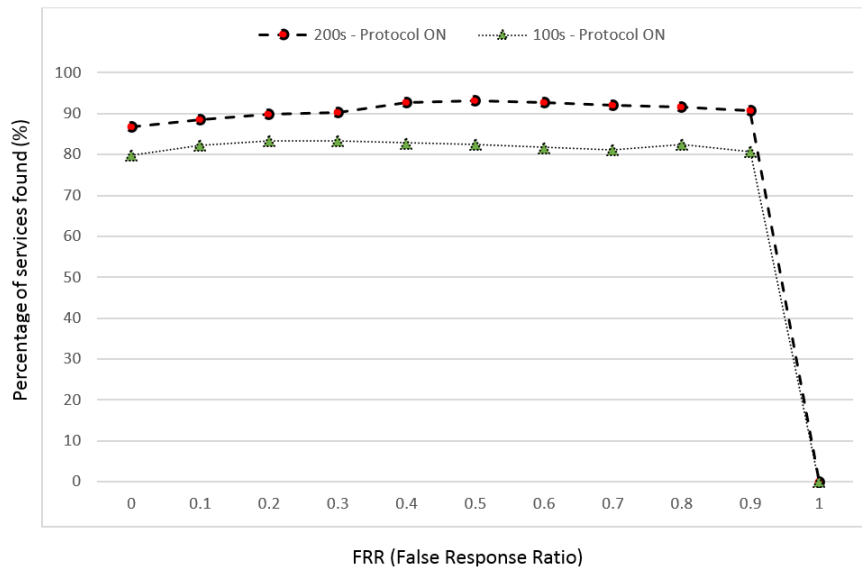


Figure 17: Invalid search response effect in service discovery protocol

The results obtained show the performance of the proposed protocol for two different mobility intervals, the percentage of services found in the network remain stable during the simulation, the totality of services in the network cannot be achieved with the current characteristics of the environment used and for the reasons mentioned before. When the FRR reaches 1, the amount of services compatible with the searches is zero because there are no services matching any search request started in the network.

Figure 18 shows the average time to find a compatibility match between services and search requests in the network. Through the use of the data type table in each nodes and the data classified through the ontology, the compatibility match and search mechanism part of the protocol should maintain the quality of searches consistent throughout the increase of FRR in the simulation. When the FRR value reaches 1 there are no services matching any of the active requests in the network, hence the action is not triggered and no time is wasted in search process.

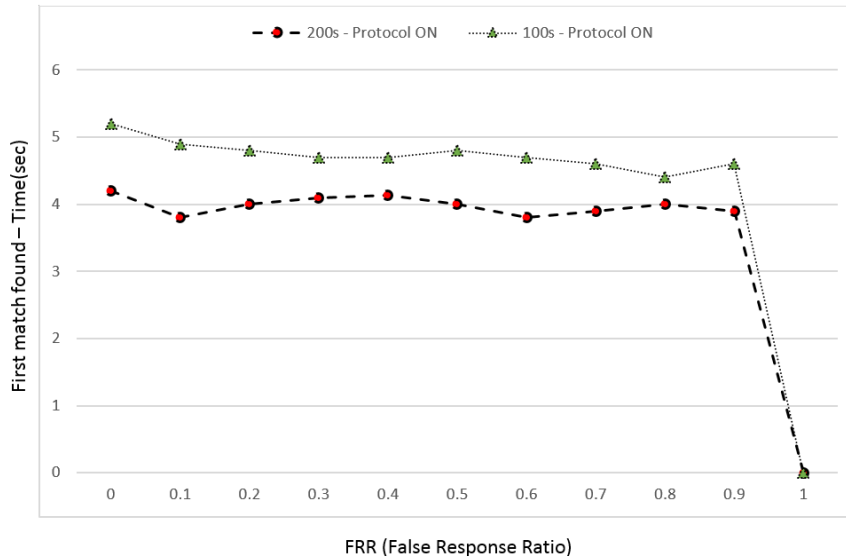


Figure 18: Invalid search response effect in compatibility match process time

Again the average time to find the first match of a service to a search request in the network remains stable with the increase of FRR value. The set of nodes with 200 seconds mobility interval perform better because of the extra time each node remains steady in the network, hence the communication path have less chance to be altered or be broken due to nodes moving or leaving the network.

Figure 19 shows the number of messages transferred in the network as a result of the message collisions and necessary re-transmission when the FRR value increases. Through the reliability in the dissemination layer explained in section 4.2.2 of the current chapter, the acknowledgement messages play an important role for the message transmission in the network, when the FRR increases the number of messages disseminated in the network decreases due to the absence of a positive response from the compatibility match of nodes in the network. When nodes receive a message, a response should be directed to the node that started the transmission, if the waiting time is higher than the set threshold, the entry for the expected response is erased, hence, when more invalid searches are placed in the network, the amount of search messages is decreased due to less possibility of message collision and the need of re-transmission of the same to invalid nodes in the network.

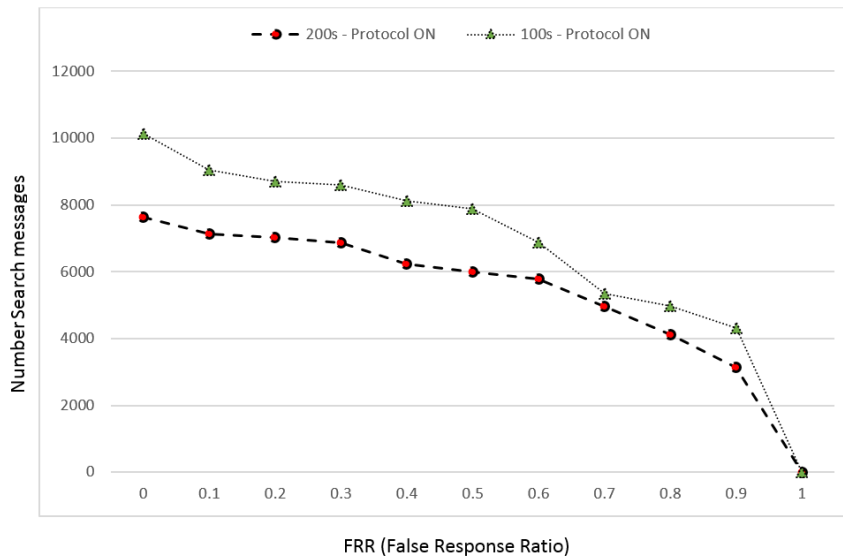


Figure 19: Invalid search response effect in number of search messages disseminated

Two different mobility intervals are used in the simulation and when the FRR equals to 1 there is no service compatible to any search request active in the network, hence the amount of search messages is zero. As shown in figure 18 the first match occurrence time remains constant although the number of search messages is decreased, as shown in figure 19, with higher values of FRR demonstrating the satisfactory performance of the protocol and the efficient adaptability to the dynamic characteristics of the network environment.

5.4 System overload with active protocol

During the dissemination, compatibility match and search process, the overload of the system is measured through the total amount of message transmitted in the network that participate in either of the mentioned process. Due to the dynamic nature of MANET networks, the protocol is constantly adapting to the changes in the topology as a response of the events created by the mobility of nodes in the network, a constant overload is experienced by each node participating in the network, measured in KB/sec.

With the increase of the FRR value, the overload of the system is reduced due to the decreasing amount of transmitted messages in the dissemination layer shown in figure 15 and search messages shown in figure 19.

Figure 20 shows the overload of the system and the effect of reduced FRR values in the system. As mentioned in section 5.3 the low possibility of message collision results in decreasing re-transmissions of messages through the network, directly visible in the necessary load carried by each node participating in the environment.

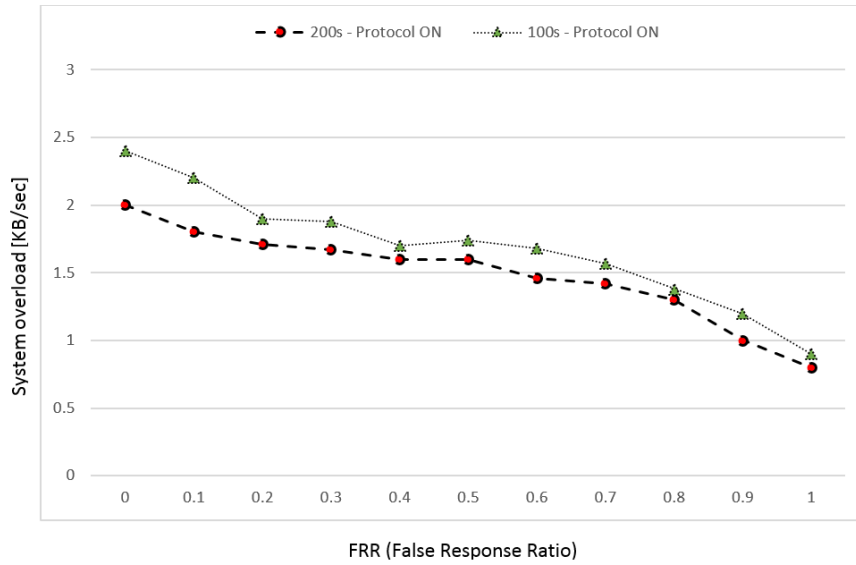
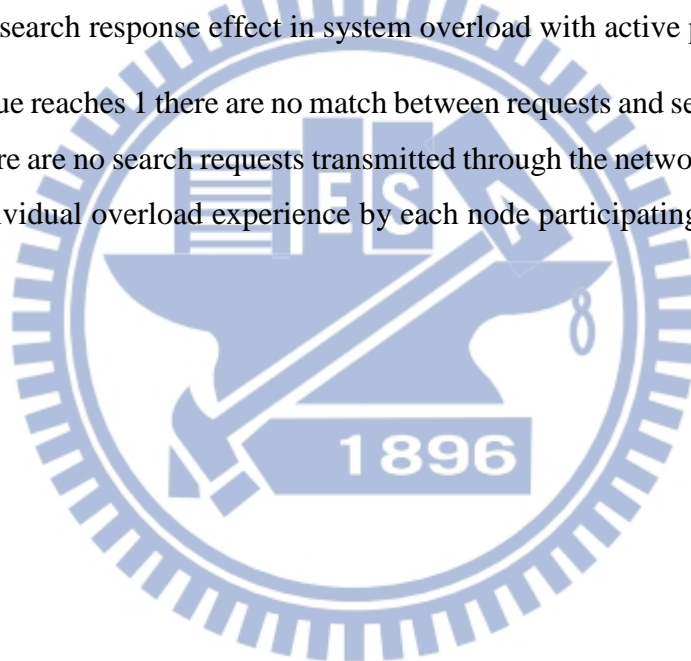


Figure 20: Invalid search response effect in system overload with active protocol

When the FRR value reaches 1 there are no match between requests and services in the network, in other words, there are no search requests transmitted through the network, essentially leaving the sum of the individual overload experience by each node participating in the network close to 0.9 KB/sec.



VI. Conclusion

Due to the different requirements and specifications to fulfill during the design of a service discovery protocol, there is no definitive set of features to claim a solution is better than others, instead the characteristics of a protocol are analyzed to fit the requirements and adequate to the attributes of a particular network as is, in this case, the MANET networks. The study of the previous works and the classification according to the different features considered in the design of the same, allowed for a detailed insight of the active problems in the field of service discovery protocol design and definition of the added values of the present thesis.

The results obtained during the simulation and testing of the proposed resource and service discovery algorithm show a high level of performance, fulfilling the expected outcomes while attaining to the limitations and specifications of MANET networks. The graphical comparisons in chapter 5 also demonstrates the robust architecture of the algorithm, executing the different phases and achieving a constant result in the amount of services discovered in the network.

6.1 Thesis contributions

- Design and development of the dissemination mechanism that makes up the dissemination layer using the ant colony optimization and infectious spread concepts to transmit messages via multiple hops and the data type classification through an ontology to create information relation and reduce the number of messages used.
- Reducing the amount of messages in the search mechanism developed using the compatibility match of data type and features of the dissemination layer, efficiently decreasing the process time for matching search requests with a proper service description available in the network.
- Development of the route composition mechanism to create and maintain the connectivity between nodes in the network, adapting to the multiple changes in the network due to nodes mobility.
- Construction of a testing environment composed of a controlled network and robust network simulator to assess the performance of the proposed resource and service discovery protocol. The number of messages transmitted in the dissemination and search phases is greatly reduced through the use of the data compatibility match and ontology classification.

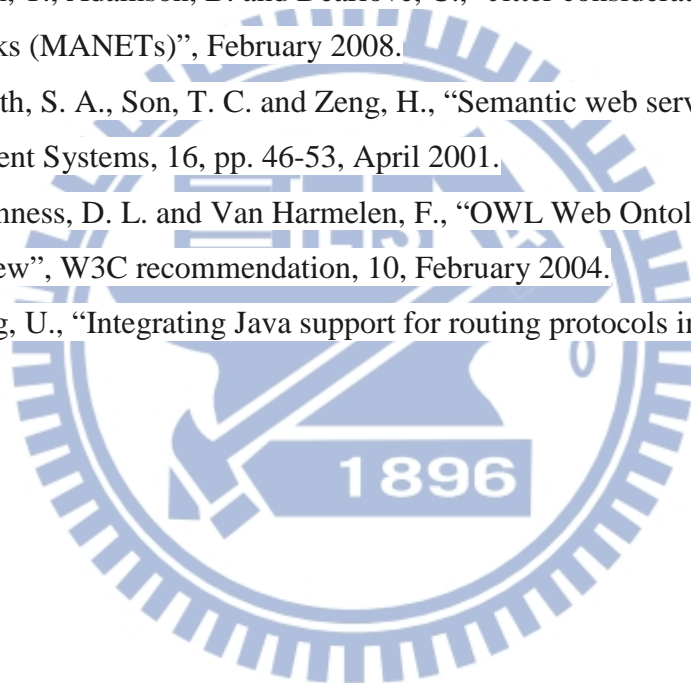
References

- [1] Weiser, M., “The computer for the 21st century”, *Scientific American*, 265, pp. 94-104, September 1991.
- [2] Masuoka, R., Parsia, B. and Labrou, Y., “Task Computing – The Semantic Web Meets Pervasive Computing”, *The Semantic Web – ISWC, Lecture Notes in Computer Science*, 2870, pp. 866-881, October 2003.
- [3] Berners-Lee, T., Hendler, J. and Lassila, O., “The Semantic Web”, *Scientific American*, 284, pp.34-43, May 2001.
- [4] Dorigo, M., Birattari, M. and Stutzle, T., “Ant colony optimization”, *Computational Intelligence Magazine, IEEE*, 1, pp. 28-39, November 2006.
- [5] Chakraborty, D., Joshi, A., Yesha, Y. and Finin, T., “Toward Distributed service discovery in pervasive computing environments”, *IEEE Transactions on Mobile Computing*, 5, pp. 97-112, February 2006.
- [6] Said, R. and Maheo, Y., “Toward a Platform for Service Discovery and Invocation in Disconnected Mobile Ad hoc Networks”, *IEEE/IFIP International Conference on Embedded and Ubiquitous Computing, EUC’08*, 1, pp. 238-244, December 2008.
- [7] Murthy, C. and Manoj, B. S., “Ad Hoc Wireless Networks: Architectures and Protocols”, Prentice Hall, March 2012.
- [8] Crow, B. P., Widjaja, I., Kim, Jeong Geun and Sakai, P. T., “IEEE 802.11 Wireless Local Area Networks”, *IEEE Communications Magazine, IEEE*, 35, pp. 116-126, September 1997.
- [9] Perkins, C. E. and Bhagwat, P., “Highly dynamic Destination-Sequenced Distance-Vector routing (DSDV) for mobile computers”, *Proceedings of the conference on Communications architectures, protocols and applications, SIGCOMM’94*, 24, pp. 234-244, October 1994.
- [10] Clausen, T., Jacquet, P., Adjih C., Laouiti, A. and Minet, P., “Optimized Link State Routing Protocol (OLSR)”, *Network Working Group*, 2003.
- [11] Perkins, C., Royer, E. and Das, S., “Ad hoc on-demand distance vector (AODV) routing”, *Internet RFC*, pp. 1-38, July 2003.
- [12] Cavin, D., Sasson, Y. and Schiper, A., “On the accuracy of MANET simulators”, *Proceedings of the second ACM international workshop on Principles of mobile computing, POMC’02*, pp. 38-43, 2002.

- [13] Andel, T.R. and Yasinsac, A., “On the credibility of manet simulations”, Computer, 39, pp. 48-54, July 2006.
- [14] Zeng, X., Bagrodia, R. and Gerla, M., “GloMoSim: a library for parallel simulation of large-scale wireless networks”, Twelfth Workshop on Parallel and Distributed Simulation, PADS 98, IEEE, pp. 154-161, May 1998.
- [15] OPNET Technologies, Inc., Acquired by riverbed on 2012, www.riverbed.com.
- [16] Chang, X., “Network simulations with OPNET”, Proceedings of the 31st conference on Winter simulation: Simulation a bridge to the future, 1, pp. 307-314, 1999.
- [17] Fall, K. and Varadhan K., “The ns manual. The VINT project. Retrieved February 10, 2011”, 2011.
- [18] Simulator, Qualnet Network, “Scalable Network Technologies” Inc., 2011, www.qualnet.com
- [19] Riley, G. F. and Henderson, T. R., “Modeling and Tools for Network Simulation”, (Klaus Wehrle, Mesut Gunes, James Gross), Springer Berlin Heidelberg, pp. 15-34, 2010.
- [20] Kurkowski, S., Navidi, W. and Camp, T., “Constructing MANET Simulation Scenarios That Meet Standards”, IEEE International Conference on Mobile Adhoc and Sensor Systems, MASS 2007, IEEE, pp. 1-9, October 2007.
- [21] Munjal, A., Camp, T. and Navidi, W.C., “Constructing rigorous MANET simulation scenarios with realistic mobility”, European Wireless Conference, IEEE, pp. 817-824, April 2010.
- [22] Camp, T., Boleng, J. and Davies V., “A Survey of Mobility Models in Wireless Ad-hoc Networks”, Wireless Communications and Mobile Computing Special Issue: Mobile Ad Hoc Networking – Research, Trends and Applications, 2, pp. 483-502, August 2002.
- [23] Ververidis, C. N. and Polyzos, G.C., “Service discovery for mobile ad hoc networks: A survey of issues and techniques”, IEEE Communications Surveys & Tutorials, 10(3), pp. 30-45, 2008.
- [24] Zhu, F., Mutka, M. W. and Ni, L.M., “Service discovery in pervasive computing environments”, IEEE Pervasive computing, 4, pp. 81-90, 2005.
- [25] Sailhan, F. and Issarny, V., “Scalable service discovery for MANET”, Third IEEE International Conference on Pervasive Computing and Communications, PerCom 2005, IEEE, pp. 235-244, March 2005.

- [26] Bloom, B.H., "Space/time trade-offs in has coding with allowable errors", Communications of the ACM, 13, pp. 422-426, July 1970.
- [27] Klein, M., Konig-Ries, B. and Obreiter, P., "Service rings – a semantic overlay for service discovery in ad hoc networks", Proceedings of the 14th International Workshop on Database and Expert Systems Applications, IEEE, pp. 180-185, September 2003.
- [28] Hwang, C., Talipov, E. and Cha, H., "Distributed geographic service discovery for mobile sensor networks", Computer Networks, Elsevier, 55, pp. 1069-1082, April 2011.
- [29] Pirrò, G., Talia D., and Trunfio, P., "A DHT-based semantic overlay network for service discovery", Future Generation Computer Systems, Elsevier, 28, pp. 689-707, April 2012.
- [30] Nidd, M., "Service discovery in DEAPspace", Personal Communications, IEEE, 8, pp. 39-45, August 2001.
- [31] Donggeon Noh, Heonshik Shin, "An Adaptive Hybrid Strategy of Integrated Service Discovery for Wireless Mobile Ad Hoc Networks", 26th IEEE International Conference on Distributed Computing Systems Workshops, ICDCS Workshops 2006, IEEE, pp. 89, July 2006.
- [32] Palmieri, F., "Scalable service discovery in ubiquitous and pervasive computing architectures: A percolation-driven approach", Future Generation Computer Systems, Elsevier, 29, pp. 693-703, March 2013.
- [33] Paolucci, M., Kawamura, T., Payne, T. R., and Sycara, K., "Semantic matching of web services capabilities", The Semantic Web – ISWC 2002, Springer Berlin Heidelberg, 2342, pp. 333-347, May 2002.
- [34] Islam, N. and Shaikh, Z. A., "Towards a robust and scalable semantic service discovery scheme for mobile ad hoc network", Pak. J. Engg. & Appl. Sci., 10, pp. 68-88, January 2012.
- [35] Varshavsky, A., Reid, B. and Lara, E., "A cross-layer approach to service discovery and selection in MANETs", IEEE International Conference on Mobile Adhoc and Sensor Systems Conference, IEEE, pp. 8-466, November 2005.
- [36] Dabrowski, C., Mills, K. and Elder, J., "Understanding consistency maintenance in service discovery architectures during communication failure", Proceedings of the 3rd International Workshop on Software and performance, WOSP'02, ACM, pp. 168-178, July 2002.

- [37] Seno, S. A. H., Budiarto, R. and Wan, T. C., “Survey And New Approach In Service Discovery And Advertisement For Mobile Ad Hoc Networks”, International Journal of Computer Science and Network Security, IJCSNS, 7, pp. 275-284, 2007.
- [38] Lenders, V., May, M. and Plattner, B., “Service discovery in mobile ad hoc networks: A field theoretic approach”, Pervasive and Mobile Computing, Elsevier, 1, pp. 343-370, September 2005.
- [39] Heissenbüttel, M., Braun, T., Wälchli, M. and Bernoulli, T., “Evaluating the limitations of and alternatives in beaconing”, Ad Hoc Networks, Elsevier, 5, pp. 558-578, July 2007.
- [40] Clausen, T., Adamson, B. and Dearlove, C., “Jitter considerations in mobile ad hoc networks (MANETs)”, February 2008.
- [41] McIlraith, S. A., Son, T. C. and Zeng, H., “Semantic web services”, IEEE Intelligent Systems, 16, pp. 46-53, April 2001.
- [42] McGuinness, D. L. and Van Harmelen, F., “OWL Web Ontology Language Overview”, W3C recommendation, 10, February 2004.
- [43] Herberg, U., “Integrating Java support for routing protocols in NS2”, 2009.



Autobiography

Emilio Sanchez, raised in Ciudad del Este, Paraguay, arrived to Taiwan in 2007 after being granted a scholarship from the government of Taiwan to follow his studies in a university of the country. He attended chinese language courses at the Chinese Language Center in National Sun Yat-Sen University in the city of Kaohsiung for a period of one year, giving him a good understanding of the chinese language. He obtained his bachelor degree in Electrical and Control Engineering department of National Chiao Tung University and started his master degree studies in the department of Electrical Engineering and Computer Science in 2012 attending the same university.

During the first academic year he met professor Huang, Ching-Yao who became his advisor and mentor, joining the wireless communication laboratory which is under the supervision of professor Huang.

The strong computer science background acquired during his high school years back in his country continued to improve throughout his bachelor degree studies and helped him gain a good understanding in circuit design and a variety of programming languages.

During the two years master studies, Emilio participated in a number of regional and international technology and business competition including Microsoft Imagine Cup 2012 obtaining the 2nd place finalist, Microsoft Project Blueprint, European Satellite Navigation Competition, APEC Intel Challenge and Advantech TiC100 Competition as finalist runner-up. Currently he is part of a start-up team working in developing a new technology to effectively pin-point moving vehicles using WI-FI signals and smartphone application based checkpoints to improve toll collection in different scenarios. His expertise areas include circuit design, fast prototyping and front-end development.

After obtaining his Master of Science degree from National Chiao Tung University, Emilio will proceed to follow his doctorate in electrical engineering and continue to work in the development of prototypes and final product for his current start-up.